

Open Research Online

The Open University's repository of research publications
and other research outputs

Functional elements of the promoter, leader and intergenic spacer regions of ribosomal RNA operon(s) of mycobacteria

Thesis

How to cite:

Ji, Y. (1993). Functional elements of the promoter, leader and intergenic spacer regions of ribosomal RNA operon(s) of mycobacteria. PhD thesis The Open University.

For guidance on citations see [FAQs](#).

© 1993 The Author

Version: Version of Record

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

**FUNCTIONAL ELEMENTS OF THE PROMOTER, LEADER
AND INTERGENIC SPACER REGIONS OF RIBOSOMAL
RNA OPERON(S) OF MYCOBACTERIA**

YUANEN JI

A thesis submitted in partial fulfilment of the
requirements of the Open University
for the degree of Doctor of Philosophy

October 1993

Laboratory of Developmental Biochemistry
National Institute for Medical Research
Mill Hill
London

Author number : P126271
Date of submission: October 1993

Dedicated to my wife Peng Jie

Library authorisation form
Form SE12 (1993)

Please return this form to the Research Degrees Office of the Open University. All students should complete Part 1. Part 2 applies only to PhD students.

Student: Yuanen JI PI: 26271

Sponsoring Establishment: National Institute for Medical Research, Mill Hill

Degree for which the thesis is submitted: Ph.D.

Thesis title: Functional elements of the promoter, leader and intergenic spacer region of ribosomal RNA operon(s) of mycobacteria.

Part 1 Open University Library Authorisation (to be completed by all students)

I confirm that I am willing for my thesis to be made available to readers by the Open University Library, and that it may be photocopied, subject to the discretion of the Librarian.

Signed: Ji Yuanen Date: 10/11/93

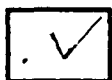
Part 2 British Library Authorisation (to be completed by PhD students only)

If you want a copy of your thesis to be held by the British Library, you must sign a British Doctoral Thesis Agreement Form and return it to the Research Degrees Office of the University together with this form. You are also required to provide the University with an unbound copy of the thesis. The British Library will use this to make their microfilm copy: it will not be returned. Information on the presentation of the thesis is given with the Agreement form.

If your thesis is part of a collaborative group project, you will need to obtain the signatures of others involved for the Agreement Form.

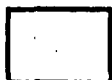
The University has decided that the lodging of your thesis at the British Library should be voluntary. Please tick either (a) or (b) below to indicate your intentions.

(a)



I am willing for the Open University to supply the British Library with a copy of my thesis

(b)



I do not wish the Open University to supply a copy of my thesis to the British Library

Signed: Ji Yuanen Date: Ji Yuanen

ABSTRACT

This study was focused on the promoter and non-coding regions of the ribosomal RNA (*rrn*) operon(s) of mycobacteria; namely, the leader and the intergenic spacer regions. Two clones containing the promoter sequences of *M.leprae* and *M.tuberculosis* *rrn* operon were sequenced, their promoter elements were identified by primer extension experiments and by comparison with *E.coli* consensus promoter sequences. Their function was tested in *E.coli* and *M.smegmatis*. The sequences of the leader and intergenic spacer regions from eight and six species respectively were established after amplification by means of PCR. Both leader and spacer regions contain antitermination elements and RNaseIII processing sites. The sequences established for these two regions also showed greater variability than the 16S rRNA gene and are suitable for phylogenetic studies. The sequences of the two *rrn* operons of *M.smegmatis* upstream from the 16S rRNA gene were cloned and sequenced. Their sequences showed that *rrnI* has a Box B element which is typical of slow-growers and that *rrnII* does not. Primer extension studies revealed that the *rrn* operon of slow-growers has a single promoter. In contrast multiple promoters were identified in the faster-growing *M.smegmatis*. Distinctive features, which are absent from slow-growers, were identified in the intergenic spacer regions of *M.smegmatis*.

INDEX OF CONTENTS

1	INTRODUCTION	1
1.1	Objectives	1
1.2	Mycobacteria	2
1.3	Molecular biology of the mycobacteria	7
1.3.1	Mycobacterial nucleic acids	7
1.3.2	Cloning and expression of mycobacterial genes	8
1.3.3	Antigenic proteins	11
1.3.4	Identification of mycobacteria by using rRNA genes	12
1.4	Gene regulation	14
1.4.1	Initiation of transcription	14
1.4.2	Regulation of mycobacterial genes	15
1.5	Prokaryotic ribosomes and ribosomal RNA	16
1.5.1	Ribosome structure and functions	16
1.5.2	Organization of rRNA operons and their transcription	18
1.5.3	Regulation of rRNA synthesis	19
1.5.4	Antitermination	21
1.5.5	Bacterial ribosomal RNA processing	23
2	MATERIALS AND METHODS	26
2.1	Bacterial strains and vectors	26
2.2	Media	26
2.3	DNA Preparation	27
2.3.1	Large scale DNA preparation	27
2.3.2	Small scale DNA preparation	29

2.3.3	Preparation of <i>M.smegmatis</i> total genomic DNA	29
2.4	RNA preparation	31
2.4.1	Preparation of <i>E.coli</i> total RNA	31
2.4.2	Preparation of <i>M.smegmatis</i> total RNA	32
2.5	Transcription start site(s) determined by primer extension	32
2.5.1	Preparation of labelled oligonucleotides	33
2.5.2	Primer extension	33
2.6	Polymerase chain reaction (PCR)	34
2.6.1	Conventional PCR	35
2.6.2	Inverse PCR	35
2.7	DNA cloning	36
2.7.1	Restriction enzyme digests and alkaline phosphatase treatment	36
2.7.2	Blunt-ending of DNA fragments	37
2.7.3	Purification of DNA fragments by agarose gel electrophoresis	37
2.7.4	Making a vector for PCR products cloning	39
2.7.5	Ligations	39
2.8	Transformation of <i>E.coli</i>	39
2.9	Electroporation of mycobacteria	40
2.10	Agarose gel electrophoresis	41
2.11	DNA sequencing	43
2.11.1	Preparation of clones for sequencing	43
2.11.2	Preparation and sequencing of double-strand DNA	43
2.11.3	Preparation and electrophoresis of sequencing gels	44

3	PROMOTER ACTIVITY OF THE <i>rrn</i> OPERONS OF <i>M.LEPRAE</i> AND <i>M.TUBERCULOSIS</i>	48
3.1	Introduction	48
3.2	Results	49
3.2.1	Cloning and sequencing of the 5'-region of the <i>M.leprae rrn</i> operon	49
3.2.2	Investigation of promoter activity in <i>E.coli</i>	49
3.2.3	Recognition of a cloning artefact	57
3.2.3.1	The method used to prepare a <i>M.leprae</i> DNA cosmid library	57
3.2.3.2	Identification of the sequence artefact upstream of the -10 Box	57
3.2.4	Expression of <i>M.tuberculosis rrn</i> in <i>E.coli</i>	60
3.2.4.1	5'-end of the <i>rrn</i> operon of <i>M.tuberculosis</i>	60
3.2.4.2	Comparison with the 5'-end of the <i>M.leprae rrn</i> operon	61
3.2.5	Expression of the <i>M.tuberculosis rrn</i> operon in <i>M.smegmatis</i>	64
3.3	Discussion	66
4	SEQUENCE ANALYSIS OF THE LEADER REGION OF THE rRNA GENE OF SLOW-GROWING MYCOBACTERIA	71
4.1	Introduction	71
4.2	Formulation of the hypothesis: the transcription start site is conserved among slow-growing mycobacteria	72
4.3	Results	73
4.3.1	Amplification and cloning of the leader region of <i>rrn</i> operons of slow-growing	73

	mycobacteria	
4.3.2	Sequencing of the leader regions of <i>rrn</i> operons of slow-growing mycobacteria	75
4.3.3	Identification of functional elements of the leader regions of slow-growing mycobacteria	77
4.3.3.1	Elements important to the start of transcription	77
4.3.3.2	Elements important to antitermination	77
4.3.3.3	Elements important to pre-rRNA processing	81
4.3.3.4	The sequence is more variable in the leader region	81
4.4	Discussion	83
5	INTERNAL TRANSCRIBED SPACER OF 16S-23S rRNA	88
5.1	Introduction	88
5.2	Results	89
5.2.1	The internal transcribed spacer of the 16S-23S rRNA spacer sequence	89
5.2.2	Identification of functional elements in the ITS	91
5.2.3	Greater sequence variability than 16S rRNA genes	94
5.3	Discussion	96
6	PUTATIVE SECONDARY STRUCTURE OF PRE-16S rRNA OF SLOW-GROWING MYCOBACTERIA AND THE INFERRED RNase PROCESSING PATHWAY	99
6.1	Introduction	99
6.2	Results and discussion	100

7	PRE-rRNA OF FAST-GROWING MYCOBACTERIA AS EXEMPLIFIED BY <i>M.SMEGMATIS</i> NCTC 8159	108
7.1	Introduction	108
7.2	Results	109
7.2.1	Cloning and sequencing of the upstream region of the first 16S rRNA gene of <i>M.smegmatis</i> NCTC 8159	109
7.2.2	Cloning and sequencing of the upstream region of the second <i>rrn</i> operon of <i>M.smegmatis</i> NCTC 8159	111
7.2.3	Comparison of the upstream sequences of the two <i>M.smegmatis</i> 16S rRNA genes	114
7.2.4	Cloning and sequencing of the ITS region between 16S and 23S rRNA genes of <i>M.smegmatis</i>	114
7.2.5	The putative secondary structure of <i>M.smegmatis</i> pre-16S rRNA and its processing	115
7.2.6	Transcription start sites of the <i>M.smegmatis rrn</i> operons	120
7.3	Discussion	122
8	GENERAL DISCUSSION	125
8.1	The organization of the mycobacterial <i>rrn</i> operon(s)	125
8.2	The presence of antitermination elements in pre-16S rRNA of mycobacteria	128
8.3	Differentiation of <i>Mycobacterium</i> species by direct sequencing of the amplified leader or ITS region	131

8.4	Expression of foreign antigens	134
8.5	Future work	135
8.5.1	Cloning the promoters of the two <i>rrn</i> operons of <i>M.smegmatis</i>	135
8.5.2	Determination of the number of pre-rRNA transcripts from each of the <i>M.smegmatis</i> <i>rrn</i> operons	137
8.5.3	Establishing sequences of the leader and ITS regions from more mycobacterial species	137
8.5.4	Identification of individual mycobacteria	137
8.6	Conclusion	139
	ACKNOWLEDGEMENTS	140
	APPENDIX A: Media	141
	APPENDIX B: Recombinant plasmids	143
	REFERENCES	145

LIST OF FIGURES

<u>Figure number and title</u>	<u>Page</u>
3.1 KEK6 plasmid construct containing the 5'-flanking region of <i>M.leprae</i> 16S rRNA gene	50
3.2 Sequencing strategy of the insert of KEK6	51
3.3 The sequence of the 5'-flanking region of the <i>M.leprae</i> 16S rRNA gene	52
3.4 Construction of pEJ106 derivatives	55
3.5 Identifying cloning artefacts by PCR	59
3.6 The sequence of the 5'-flanking region of <i>M.tuberculosis</i> 16S rRNA gene	62
3.7 Transcription start site of <i>M.tuberculosis</i> <i>rrn</i> operon in <i>E.coli</i>	63
3.8 Transcription start site of <i>M.tuberculosis</i> <i>rrn</i> operon in <i>M.smegmatis</i>	67
4.1 Amplification of the leader regions of mycobacteria	74
4.2 The alignment of leader sequences of slow-growing mycobacteria	79
5.1 Amplification of the ITS region of mycobacteria	90
5.2 Alignment of the ITS sequences of slow-growing mycobacteria	93
6.1 The secondary structure of <i>M.leprae</i> pre-16S rRNA	103
6.2 The secondary structure of <i>M.tuberculosis</i> pre-16S rRNA	105
7.1 The sequences upstream from the 16S rRNA genes of <i>M.smegmatis</i>	110

7.2	<i>rrnI</i> operon of <i>M.smegmatis</i> contains a slow-grower-like Box B element	112
7.3	The strategy of inverse PCR	113
7.4	The sequence of <i>M.smegmatis</i> ITS region	116
7.5	Comparison of helix S4 of <i>M.smegmatis</i> with that of slow-growers	118
7.6	RNaseIII processing sites of <i>M.smegmatis</i> pre-16S rRNA	119
7.7	Transcription start sites of <i>M.smegmatis</i> <i>rrn</i> operons	121

LIST OF TABLES

<u>Table number and title</u>	<u>Page</u>
3.1 The promoters of <i>M.tuberculosis</i> and <i>M.leprae</i> <i>rrn</i> operons function in <i>E.coli</i>	56
3.2 Comparison of <i>rrn</i> operon promoter sequences of <i>M.leprae</i> , <i>M.tuberculosis</i> , <i>B.subtilis</i> , and <i>E.coli</i>	65
4.1 The sequences of helix10 of the V2 regions confirm the species studied	76
4.2 Comparison of properties of the leader regions of mycobacteria	82
4.3 Comparison of invariant leader regions of mycobacterial <i>rrn</i> operons with those of other species	85
5.1 Comparison of properties of the ITS regions of mycobacteria	95

CHAPTER ONE

Introduction

1.1) Objectives

Slow-growing mycobacteria have a single ribosomal RNA (*rrn*) operon (Bercovier et al., 1986). The organization of the three rRNA genes into one operon is one way of ensuring that mature 16S rRNA, 23S rRNA and 5S rRNA are synthesised in equimolar amounts. This strategy requires that the transcription process be highly efficient; that is, a complete copy of the operon should be produced each time transcription is initiated, and the transcript should be efficiently processed. The objectives of this project were to study the structure of *rrn* operon(s) in mycobacteria; in particular to identify the promoter and the antitermination elements involved in the transcription of the single *rrn* operon of slow-growing mycobacteria from the start to finish without interruption. The establishment of the sequences for the flanking regions of the mature rRNA genes should enable one to build a general secondary structure for the pre-rRNA molecules, from which a scheme of processing mechanism can be inferred. The *rrn* operon is of interest for two other reasons. Firstly, there is a practical need for identifying closely related mycobacterial species. The flanking regions of *rrn* operons are the more likely sites for species specific sequences than 16S rRNA genes, because they are presumably under less evolutionary pressure to maintain a

particular three dimensional structure. Secondly, it is clear that a large proportion of mycobacterial promoters are not active in *E.coli* (Dale and Patki, 1990), and do not possess the typical *E.coli* consensus promoter sequence; therefore the identification of the promoter for the *rrn* operon is of particular interest. If the promoter does function in *E.coli*, then it can be used to drive the mycobacterial genes in *E.coli* and allow high level production of proteins of interest. The proteins produced in this way might prove of great value in the study of protective immunity, allowing the study of particular "protective" antigens or individual determinants of pathogenicity.

A representative fast-growing mycobacterium, *M.smegmatis*, was also studied. Fast-growers have two *rrn* operons compared with the single *rrn* operon of the slow-growers (Bercovier et al., 1986). The aims of the study of *M.smegmatis* were to clone and establish sequences for the leader regions of the two *rrn* operons, and the internal transcribed spacer region, in order to identify features of the fast-growing mycobacterial *rrn* operons which are different from those of slow-growing mycobacteria.

1.2) Mycobacteria

The generic name *Mycobacterium*, originally established by Lehmann and Neumann (Lehmann and Neumann, 1896), encompasses over fifty approved species, including the pathogenic tubercle and leprosy bacilli, with new species

still being discovered. Mycobacteria belong to the high guanine-plus-cytosine subdivision of Gram-positive bacteria. The G+C content of their DNA ranges from 62-70% (Goodfellow & Wayne, 1982) with the exception of that of *M. leprae*, which is 56% (Clark-Curtiss et al., 1985). Mycobacteria can be described as being aerobic acid-alcohol-fast actinomycetes that usually grow as slightly curved or straight rods. All have *meso*-diaminopimelic acid, arabinose and galactose in their walls, produce straight chain fatty acids and the characteristic long-chain mycolic acids (Goodfellow & Wayne, 1982). The high proportion of lipids renders the organisms difficult to stain by dyes, requiring prolonged exposure, frequently with the application of heat. When stained however, they are resistant to decolorisation by acid. Consequently they have been termed acid-fast bacteria, and this property has formed the basis of specialised stains for detection of the organisms in biological samples (Fite, 1947). Mycobacteria can be classified into two major categories, depending on their growth characteristics in *in vitro* culture systems. Most slow growers take ten days or longer to yield visible colonies (except *M. leprae* which has not been grown *in vitro*), while the fast growers require only two to four days. The optimum temperatures for the growth of species in either of these categories range from 28-45°C. Taxonomically the slow and fast growers can be further subdivided into "complexes" which comprise closely related species of mycobacteria.

Mycobacteria were among the first bacterial pathogens

of man to be studied. Hansen (1874) identified the leprosy bacillus, *M.leprae*, in the lesions of leprosy patients in 1873 and Koch (1882) clearly established that *M.tuberculosis* was the causative agent of tuberculosis (TB). Today, in spite of significant progress in diagnosis, treatment and prevention, TB and leprosy still remain major public health problems worldwide.

Leprosy is a chronic mycobacterial disease caused by the obligate intracellular parasite, *M.leprae*, which afflicts 5.5 million people in the world today (Noordeen et al., 1992). The causative agent remains one of the few major pathogens not grown successfully in culture. Leprosy comprises a clinical and immunological spectrum ranging from tuberculoid types of leprosy, where patients have a high level of cell-mediated immunity, to lepromatous leprosy, where the selective immunological unresponsiveness of the patients to antigens of *M.leprae* enables the organism to multiply intracellularly virtually unchecked, especially in the macrophages and Schwann cells. In both cases, nerve damage results in deformity and subsequent social stigmatization. Effective chemotherapy exists, but a preventive vaccine would be preferable, since the disease is common in areas remote from medical services. Although a specific vaccine against leprosy does not exist, trials with BCG indicate that this mycobacterium confers some protection against leprosy.

Despite major advances in diagnostic techniques and effectiveness of treatment, *M.tuberculosis* remains the

bacterial pathogen causing the greatest amount of chronic disease and death throughout the world. More than 20% of the world's population is infected with the tubercle bacillus. Every year there are eight to ten million new cases and three to five million deaths attributable to tuberculosis (Kochi, 1991).

M.tuberculosis and *M.leprae* are closely related and show several important biological similarities. They both grow inside host cells for long periods of time, and both infections are controlled by cell-mediated mechanisms, rather than by antibody responses. In addition, both species show the phenomenon of persistence, where bacilli in tissues can survive for long periods of time in the presence of bactericidal concentrations of anti-tuberculous or anti-leprosy drugs. The obvious similarities between *M.tuberculosis* and *M.leprae* mean that research on one of these organisms is often applicable to the other.

BCG, an attenuated strain of *M.bovis*, is the most widely used vaccine in the world and is also a potentially useful vaccine vehicle for delivering protective antigens of multiple pathogens. This vaccine has powerful adjuvant activity, being able to stimulate both the humoral and cell-mediated immune systems with a low rate of serious complications. The ability of BCG to protect against tuberculosis has varied considerably in a number of human vaccine trials. For example, BCG has been shown to have poor efficacy in southern India and the southern United States, whilst it is effective in the United Kingdom and East Africa

(reviewed by Fine and Rodrigues, 1990). The high incidence of tuberculosis infections around the world and the inability of BCG to protect certain populations clearly shows that a better vaccine is required. An improved BCG vaccine might also provide greater protection against leprosy.

Organisms of the *M.avium* complex are opportunistic pathogens common in the environment and are of renewed interest today because significant proportions of individuals with acquired immune deficiency syndrome (AIDS) have disseminated pulmonary infection with organisms of the *M.avium-M.intracellulare* complex (Good, 1985; Collins, 1988). *M.paratuberculosis* is a very closely related organism which is the causative agent of Johne's disease (or paratuberculosis) in cattle and sheep and has also been implicated in the etiology of human Crohn's disease (Chiodini, 1989; Elsaghier et al., 1992), although this conclusion remains controversial. Infection with *M.marinum* is associated with aquatic activity and is the cause of the "swimming pool granuloma" and the "fish fanciers finger". There is universal agreement that the infection occurs as a result of contamination of trivial skin lesions. *M.ulcerans* causes progressive malignant ulcers of skin and deeper tissues in several tropical regions such as Nigeria, Zaire, Malaysia, New Guinea, Mexico and other countries. The variety of fast growing environmental mycobacteria may also manifest as post-surgical and post-traumatic infections in immuno-compromised patients. Few vertebrates are immune to

mycobacterial infections and some individual species, such as *M.bovis* and *M.avium*, not only infect several animal species but may be transmitted between species.

Slow-growing mycobacteria therefore form a large and important group of human and animal pathogens. Despite the early discovery of these organisms, the slow growth rate of pathogens such as *M.tuberculosis* and the failure of *M.leprae* to grow *in vitro* have constrained the study of the basic biology and host-pathogen relationship of these bacteria. Use of recombinant DNA methodology has great potential for increasing our understanding of the basic physiology of these organisms and also for deciphering the mechanisms of pathogenicity, for the production of reagents for diagnosis, for designing effective new drugs and thus eventually for prevention of TB and leprosy altogether.

1.3) Molecular biology of the mycobacteria

Although major and fundamental breakthroughs have been made in relation to the understanding of basic genetic mechanisms in bacteria, especially in *E.coli*, these aspects have not been well studied in mycobacteria. The relatively (and in many cases extremely) slow growth rate of mycobacteria do not make them choice material for the geneticist to study. Also, for a long time, systems for exchange of genetic material between strains have not been as easy to establish in this genus as in several other genera.

1.3.1) Mycobacterial nucleic acids

Among the eubacteria, the mean guanine and cytosine (G+C) content shows wide variation, ranging from 25% to 75%. Mycobacteria thus belong to the high G+C subdivision of the Gram-positive bacteria, which also include such actinomycetes as *Streptomyces*. *Streptomyces* have about 72 mole per cent G+C in their DNA, *M.tuberculosis* 63-65% and *M.leprae* 56% (Imaeda et al., 1982; Jacobs et al., 1986).

The size of the bacterial genome is measured by optical reassociation studies. Estimates from mycobacterial strains in one such study by Bradley (1973) gave 2.5 to 4.5×10^9 daltons for 15 different mycobacterial species. The G+C ratios in mycobacteria from this and various other studies have given an average range of 60 to 70% with the exception of *M.leprae* (reviewed by Clark-Curtiss, 1990).

Modification of DNA plays an important part in permitting the cell to distinguish between its "own" and "foreign" DNA. For example, nucleotide methylation protects the DNA of a cell from degradation by endogenous restriction endonucleases. Modified bases such as 6-methylaminopurine and 5-methylcytosine (Grange, 1982; Hottat et al., 1988) have been detected in mycobacteria. The report of the existence of the restriction endonuclease system in *M.tuberculosis* (Rado et al., 1976), confirms the importance of base modification in mycobacteria.

1.3.2) Cloning and expression of mycobacterial genes

Over the past seven years the genomes of several mycobacterial pathogens, including *M.leprae* and *M.tuberculosis*, have been cloned in *E.coli* using both

plasmid and phage expression vectors (Clark-Curtiss et al., 1985; Young et al., 1985a,b; Jacobs et al., 1986). Young et al. (1985a,b) constructed libraries of *M.leprae* and *M.tuberculosis* genomic DNA sequences in the λ gt11 expression vector. Clark-Curtiss et al., (1985) constructed genomic libraries using cosmid and expression vectors in *E.coli* and demonstrated that mycobacterial DNA could be expressed in *E.coli* under the control of an *E.coli* promoter; mycobacterial promoters were mostly poorly or not at all recognised by the *E.coli* transcription machinery. Only a small number of BCG promoters are known to be active in *E.coli* (Thole et al., 1985), although a few are strongly active (Sirakova et al., 1989). No data are available about the real number of mycobacterial promoters which are active in *E.coli* and the nature of the gene products whose synthesis they control.

The first cloning host used was *E.coli* because of the wealth of genetic knowledge and procedures for gene cloning available for this organism. *E.coli* is a suitable host for the primary cloning of many mycobacterial genes. Aside from the inability of *E.coli* to utilise mycobacterial promoters efficiently, an inherent limitation in using *E.coli* as a cloning host is that certain products which require complex biosynthetic pathways may not be formed. Genes which may be missed by such an approach, include those involved in formation of antigenic polysaccharides and lipids, such as the phenolic glycolipid of *M.leprae* and the glycopeptidolipids of *M.avium*, which are produced by

incompletely understood biosynthetic pathways catalysed by several gene products. Multiple genes would need to be cloned in order to detect such products. In addition to *E.coli*, *Streptomyces lividans* has been explored as a host capable of expressing many mycobacterial genes, or perhaps groups of linked genes coding for a biosynthetic pathway or a complex antigen, which are initiated by mycobacterial expression signals.

Readily cultivable mycobacteria such as *M.smegmatis* or *M.bovis* BCG should be ideal surrogate hosts for the faithful expression of genes from pathogenic mycobacteria. Any use of a mycobacterial strain as a host for recombinant DNA requires the development of an efficient transformation system in combination with versatile cloning vectors. The development of phage (Jacobs et al., 1987), plasmid (Snapper et al., 1988) and gene replacement (Husson et al., 1990) systems for the introduction of recombinant DNA into mycobacteria has allowed the development of mycobacteria as cloning hosts. The use of *M.smegmatis* as a cloning host has been greatly facilitated by the isolation of a mutant (designated *M.smegmatis* mc²155), that can be transformed at much higher efficiencies than the parental strain using electroporation (Snapper et al., 1990). Another important requirement in developing cloning hosts is the absence of host-dependent modification and restriction enzyme systems. *M.tuberculosis* has been shown to possess such a system (Rado et al., 1976), *M.avium* has a plasmid-mediated modification and restriction system (Crawford et al., 1981), and

restriction endonucleases have been isolated from *M.habana* and *M.fortuitum* (Shankar & Tyagi, 1992a,b). However, *M.smegmatis* mc²155 has been shown to lack such a system (Snapper et al., 1990).

1.3.3) Antigenic proteins

Certain mycobacterial proteins have been described as "major antigens" based on their ability to evoke a strong humoral immune response. The major focus of recent mycobacterial research is to identify and characterize individual proteins of mycobacteria which are involved in interactions with the immune system. A group of mycobacterial proteins which have been studied in detail are the so called heat shock proteins (hsps), which enable cells to survive environmental stresses such as elevated temperature and oxidative stress. Heat shock proteins appear to be major immunogens in many infections, and it has been proposed that hsps provide a link between immunity to infection and autoimmune disease. They can be classified according to molecular weight into families, termed hsp90, hsp70, hsp60 and small hsps (Lindquist & Craig, 1988); their functions are associated with and influence other proteins (Ellis, 1987). In mycobacteria, several cloned antigens have been classified as heat shock proteins: the *M.tuberculosis* 70 kDa (kilodalton) antigen (Young et al., 1988) and its 70 kDa *M.leprae* homologue (Garsia et al., 1989) belong to the hsp70 family; the 65 kDa antigen of *M.leprae* (Young et al., 1988) and *M.tuberculosis* belongs to the hsp60 family; the 18 kDa antigen of *M.leprae* was thought to be unique in being

the only bacterial member of a group of loosely related heat shock proteins similar to alpha-crystallin in eukaryotic cells (Nerland et al., 1988), now it also has been found in *M.avium* and *M.simiae*; the 12 kDa antigen of *M.tuberculosis* is a homologue of the *groES* gene product of *E.coli* (Baird et al., 1988), and together with the *groEL* gene product, it is involved in the heat shock response in *E.coli*. These proteins are among the most highly conserved proteins in existence and the proteins themselves have normally been found to have more than 50% sequence homology in organisms ranging from *E.coli* to man.

Another important group of mycobacterial protein antigens are secreted proteins which would be expected to be protective antigens responsible for rapid recognition of bacilli by host lymphocytes (Abou-Zeid et al., 1988). Up to now the identified mycobacterial secreted proteins include *M.bovis* BCG α -antigen (Matsuo et al., 1988), *M.tuberculosis* 32 kDa antigen (Borremans et al., 1989) and most important, *M.bovis* BCG antigen MPB64 (Yamaguchi et al., 1989). All these proteins possess a signal peptide associated with these secreted mycobacterial proteins.

1.3.4) Identification of mycobacteria by using rRNA genes

The classical methods for the identification of mycobacteria rely primarily on phenotypic, chemotaxonomic and serotaxonomic tests performed on cultures. To identify slow-growing mycobacteria is a time-consuming procedure, for example it can take as long as 10 weeks to diagnose tuberculosis based on the phenotypic characters of

M.tuberculosis. Various methods including immunological techniques, DNA homology and related analyses (Baess, 1979; 1983; McFadden et al., 1987; Gross & Wayne, 1970; Imaeda et al., 1988), have been tested for the rapid diagnosis of mycobacterial infections and proved useful in determining taxonomic relationships. Nevertheless reliable identification of mycobacteria remains problematical.

A correlation between the growth rate of mycobacteria and the number of rRNA genes they possess has been observed; fast-growers possess two sets of rRNA genes and slow-growers possesses only one set (Bercovier et al., 1986). rRNAs, especially 16S rRNA, are nowadays generally used for characterising the phylogenetic relationship among closely related species. rRNA is an essential constituent of bacterial and eukaryotic ribosomes (Fox et al., 1980; Woese et al., 1983), and the 16S rRNA molecule is functionally constrained such that it is highly conserved, with rare sequence changes in certain positions, However, the location of these changes is specific to the group or species in which they occur (Gray et al., 1984; Stackebrandt & Woese, 1981; Woese, 1987; Dams et al., 1988). The specificity of the 16S rRNA sequence and the information content of the molecule, is sufficient to allow both statistically valid phylogenetic analysis and species identification, once the sequence has been determined.

Rogall et al. (1990a,b) has successfully established 16S rRNA sequences for more than twenty mycobacterial species by direct sequencing of PCR amplified DNA fragments,

thus identifying a highly variable region. Based on the variable region, a few closely related species such as *M.avium* and *M.intracellulare* can be distinguished. Moreover, *M.leprae* 16S rRNA gene contains a unique sequence close to the 5'-end which was used to develop DNA probes and a PCR test for the identification of *M.leprae* in tissue samples (Cox et al., 1991; Teske et al., 1991).

1.4.) Gene regulation

1.4.1) Initiation of transcription

Initiation requires the core RNA polymerase, which is composed of two α chains, one β chain and one β' (Chamberlain, 1982), σ factor, the appropriate nucleoside triphosphate (always a purine), and a special promoter. Promoters differ in their sequences and their strength of polymerase binding. Promoters have similar but not identical sequences in two regions, -35 and -10 with respect to initiation of transcription at +1. The consensus sequence at -35 is TTGACA, while that at -10 is TATAAT (Harley & Reynolds, 1987), whose AT pairs (weaker than GC) promote the required melting. Initiation starts with the σ factor searching and binding directly to the promoter sequences (McClure, 1985; Helmann & Chamberlain, 1988). After transcription is initiated, the σ factor is then replaced by the *nusA* protein, marking the transition from the initiation to the elongation phase. An important regulatory mechanism is a shift in the σ factors in the cell: the elements that select the set of transcription start sites (promoters)

recognized by each RNA polymerase.

The analyses have suggested that the majority of *E.coli* promoters fall into three basic categories; those recognized by $E\sigma^{70}$, those recognized by $E\sigma^{32}$ and those promoters recognized by $E\sigma^{54}$ (Collado-Vibes et al., 1991; Gralla, 1990). In *B.subtilis*, a Gram positive bacterium, nine types of σ factors that are used for gene expression at different stages have been reported (Gitt et al., 1985; Fujita et al., 1989).

The transcription initiation is also regulated by regulatory proteins. They can be either repressors, for example the repressor protein for the *lac* operon, or activators like the AraC protein that binds to *araC* gene of *BAD* operon to activate the operon in the presence of arabinose, and catabolite activator protein (CAP). CAP, in the presence of cAMP, can bind to the *lac* promoter causing the DNA to bend considerably. This bending somehow activates transcription, possibly by exposing a strong binding site for RNA polymerase (Gartenberg & Crothers, 1988; Liu-Johnson et al., 1986)

1.4.2) Regulation of mycobacterial genes

When this study started more than three years ago, only a few mycobacterial promoters appeared to be active in *E.coli*, namely those of heat shock protein genes. Since then, two promoters of the 16S rRNA gene from mycobacteria have been found to be active in *E.coli*. They are the *M.leprae rrn* promoter (Sela & Curtiss, 1991) and *M.tuberculosis rrn* promoter (Ji et al., 1994, the study is

illustrated in Chapter 3). There is little information about mycobacterial promoters other than those of the *rrn* operon. The main difficulty is to extract full length mRNA transcripts from mycobacterial cells. Recently efficient methods for the extraction of mRNA from mycobacteria have been developed (Kinger et al., 1993; Patel et al., 1991), which will certainly benefit mycobacterial gene promoter study.

1.5) Prokaryotic ribosomes and ribosomal RNA (rRNA)

1.5.1) Ribosome structure and functions

The ribosome is an essential component of the protein-synthesizing machinery of prokaryotes. The *E.coli* ribosome comprises approximately 60% RNA and 40% protein. The small ribosomal subunit contains one 16S rRNA molecule together with single copies of twenty-one different ribosomal proteins; the large ribosomal subunit contains 5S and 23S rRNAs together with thirty-one proteins.

The size for 16S rRNA of prokaryotes is about 1500 nucleotides, with several nucleotides being modified. For example, in *E.coli* ten of the 1542 nucleotides of 16S rRNA are modified by methylation (Arnstein & Cox. 1992). The 16S rRNA component of the small ribosomal subunit plays an active role in the initiation, elongation, and termination steps of protein biosynthesis (Dahlberg, 1989). For example, the anti-Shine-Dalgarno sequence participates in binding mRNA by base pairing with the Shine-Dalgarno sequence (Shine & Dalgarno, 1974) located near to, but upstream from, the

AUG codon of the mRNA. This interaction is probably also important for maintaining mRNA in the correct reading frame (Weiss et al., 1987).

The 23S rRNA and 5S rRNA components comprise about 3000 and 120 nucleotides respectively. It has been established that 23S rRNA contains important functional regions such as elongation factor domains and the peptidyltransferase domain (Raué et al., 1990). Although 5S rRNA is essential for ribosome activity, its role is still unknown. It is likely, however, to be involved in an interaction with 16S rRNA during ribosome function, by analogy with known interactions of 5S rRNA and 18S rRNA in eukaryotes (Azad, 1979).

The models of common secondary structure for the rRNAs were primarily based on comparisons of the sequences of rRNAs of different phylogenetic origins. Evidence was obtained from situations where there is poor conservation of primary structure in a feature of proposed secondary structure, which is preserved because of compensatory base changes (e.g. an A→G change in one strand of the RNA is compensated for by a U→C change in the complementary base of the other strand).

Based on the above strategy, Noller (1984) proposed the secondary structures for 16S rRNA, 23S rRNA and 5S rRNA. These secondary structures serve as models for all other homologous rRNA species.

A large database comprising the primary structures of 16S rRNA genes of bacteria has been established. The data show that certain features of the gene are more highly

conserved than others. It is thus possible to use less conserved regions in 16S rRNA gene for phylogenetic studies. Comparison of 16S rRNA sequences is now a widely accepted method for establishing phylogenetic relationships (Woese, 1987).

1.5.2) Organization of rRNA operons and their transcription

In prokaryotes the number of rRNA operons (*rrn*) varies from species to species. There are seven *rrn* operons in *E.coli* (Brosius et al., 1981), ten in *B.subtilis* (Ogasawara et al., 1983), but only one in slow-growing mycobacteria (Bercovier et al., 1986). Each operon contains the genes for the three ribosomal RNAs in the order 16S, 23S and 5S, together with one or more genes for tRNA. Each gene is separated from the next by a transcribed spacer. The number and location of the tRNA genes varies. In four of the seven operons of *E.coli* the spacer between the 16S and 23S rRNA genes contains a tRNA^{Glu} gene, whereas in the remaining three operons the spacer contains genes for tRNA^{Ile} and tRNA^{Ala} (Lund & Dahlberg, 1976; Lund & Dahlberg, 1977; Young et al., 1979; Brosius et al., 1981). However, eight of ten *rrn* operons in *B.subtilis* do not bear tRNA genes (Loughney et al., 1982). There may also be tRNA genes following the 5S rRNA. In *E.coli* the operon *rrnD* contains two 5S rRNA genes with a tRNA^{Thr} gene between them (Duester & Holmes, 1980).

In *E.coli*, each *rrn* operon contains tandem rRNA promoters, called P1 and P2, separated by about 120 bases (Brosius et al., 1981; Csórdas-Tóth et al., 1979; deBoer et al., 1979; Glaser & Cashel, 1979; Shen et al., 1982; Young

et al., 1979). Tandem promoters also have been found in *B.subtilis* *rrn* operons (Ogasawara et al., 1983; Stewart & Bott, 1983). The *E.coli* promoter consensus sequences (Harley & Reynolds, 1987) can be found in both P1 and P2. A strikingly conserved sequence surrounding the -10 box of P1 (Morgan, 1982; deBoer et al., 1979) was identified and is thought to play an important role in the regulation of *rrn* operon expression (Travers, 1980a).

1.5.3) Regulation of rRNA synthesis

(i) Stringent control (for review see Nomura et al., 1984): The stringent control of wild type bacteria is to shut down protein synthesis rapidly when starved of amino acid, an event followed quickly (within a few minutes) by a dramatic down-regulation of the transcription of genes coding for 'stable' RNA (rRNA and tRNA). This downregulation is not reflected in the mRNA population in general. It is characterized by an increased synthesis of guanosine tetraphosphate (ppGpp) (Gallant, 1979). This molecule is synthesized by a ribosome-associated enzyme, the product of the *relA* gene, when the appropriate uncharged tRNA is bound at the A site of the ribosome. GTP or GDP is phosphorylated by a molecule of ATP: $GTP(or\ GDP) + ATP \rightarrow ppGpp + AMP$ in the absence of the amino acid substrate for normal protein synthesis. The function of ppGpp is thought to be to directly regulate transcription initiation from stable rRNA promoters by interacting with RNA polymerase. Travers et al. (1982) have proposed that ppGpp alters the equilibrium between different classes of RNA polymerases, only one class

of which efficiently triggers initiation at rRNA and tRNA promoters. It has been shown that only P1 of *rrn* operon promoters of *E.coli*, and not P2, is subject to stringent control (Sarmientos et al., 1983). The highly GC-rich sequences are found at the end of the -10 box of the stringently controlled promoters. Mutation in this region leads to loss of ppGpp sensitivity *in vitro* (Mizushima-Sugano & Kaziro, 1985; Travers, 1980b) and *in vivo* (Lamond & Travers, 1985). Since such GC-rich sequences can be identified in P2 promoters, additional sequence determinants must also be required for the function of stringent control (Lamond & Travers, 1985).

(ii) Growth rate-dependent control: Under most growth conditions, the cellular concentration of ribosomes is proportional to the growth rate (i.e. to the total protein synthesis rate) (for review see Nomura et al., 1984). Such a regulation in which the level of total protein synthesis required for growth is achieved by controlling the number of ribosomes, rather than modulating the activity of individual ribosomes, is called growth rate-dependent control. The molecular mechanisms responsible for growth rate control are not clear yet. One possibility is that the regulation is a consequence of stringent control by ppGpp, because it has been observed that the ppGpp concentration decreases with increasing growth rate (Ryals et al., 1982ab). The study of the *tyrT* operon suggesting that the stringent discriminator region is essential for growth rate regulation also implicates ppGpp (Travers et al., 1986). Another possibility

is a feedback control mechanism as suggested by Nomura and coworkers (Gourse et al., 1986; Nomura et al., 1984). Their results suggest that over production of a full complement of functional rRNA molecules leads to feedback inhibition of rRNA synthesis. They further noticed that only transcription from the P1 promoter is subject to feedback control (Gourse et al., 1986).

1.5.4) Antitermination

In prokaryotic cells, ribosomes start to translate an mRNA before its synthesis is completed. This kind of coupling of transcription and translation is thought, in a way not yet clear, to prevent transcription termination at termination sites within protein-coding regions. Since *rrn* operons can be only transcribed and not translated, it is likely that an antitermination mechanism exists which takes the place of the coupling of transcription and translation in reducing premature termination. The facts that insertion of Tn9, Tn10 and IS1 containing strong Rho-dependent transcriptional terminators downstream of the promoter and 5' leader region, cause far less transcriptional polarity in *rrn* operons than protein-coding operons (Brewster & Morgan, 1981; Morgan, 1980), confirm the existence of an antitermination mechanism in *rrn* operon. By using operon fusion assays, it was determined that neither RNA polymerase modification nor ribosomal protein can cause reduced termination within *rrn* operons (Holben & Morgan, 1984; Holben et al., 1985); the assays also showed antitermination sequences are localized in the leader region of *rrn* operons.

Antitermination is well established in the regulation of transcription from the major promoters of bacteriophage lambda. The leader region of *rrn* operons contain sequences homologous to the so-called Box A, B and C sequences in the *nut* region of lambda, which are required for antitermination (Holben et al., 1985; Li et al., 1984). However, the arrangement of these sequences is not the same in *rrn* operons as in lambda. In lambda, Box B is 3' to Box A; in the *rrn* anti-terminator sequence it is 5' to Box A (Berg et al., 1989). Berg et al. (1989) also carried out mutation analysis of the Box sequences and indicated that (1) Box A was sufficient for terminator readthrough; (2) there was no positive requirement for Box B element. They noticed, however, that mutation in the box B stem region could decrease or eliminate terminator readthrough even in the presence of Box A. The Box A and Box B elements also exist in *E.coli rrn* 16S-23S spacer regions in the same order as they are in the leader region (Young et al., 1979; Morgan, 1986; Harvey et al., 1988), suggesting that during transcription the antitermination mechanism needs to be renewed at certain intervals.

Other genetic and biochemical experiments have also suggested that some of the factors required for transcription antitermination in bacteriophage lambda are involved in the *rrn* system. These factors include NusA, NusB, NusG and ribosomal protein S10 (Li et al., 1992; Mason et al., 1992). Recently Nodwell and Greenblatt (1993) demonstrated that a complex of the transcription

antitermination factors NusB and ribosomal protein S10, interact specifically with Box A RNA, but neither factor can bind Box A alone. Furthermore, mutations that prevent antitermination activity *in vivo* of the ribosomal Box A also impair binding by NusB-S10. The involvement of S10 protein in the antitermination mechanism is thought to be crucial to the efficient synthesis of ribosomal RNAs. Only in the presence of free S10 protein can the cell make full-length rRNA transcripts, thus tightening the regulation of the synthesis of the ribosomes.

1.5.5) Bacterial ribosomal RNA processing

In bacteria the complex *rrn* operon is transcribed into long precursor molecules that include the spacer elements. Only 1-2% of rRNA is present as a long precursor (King & Schlessinger, 1983), presumably because it is efficiently processed by endonucleases while it is still being transcribed. The processing is multistep (Schlessinger, 1980): (a) primary processing leads to formation of intermediate precursors of rRNA from the long transcripts, and (b) maturation steps produce mature termini of rRNA from precursors. A critical difference between the step (a) and (b) is that the first cleavage occurs *in vitro* with naked pre-rRNA or with pre-ribosomes as substrate, but final maturation depends on the pre-formation of the complex of rRNA and r-proteins. In studies to date, three enzymes have been clearly implicated in the processing of the nascent rRNA transcript. They are RNaseIII, RNaseP and RNaseE (Gegenheimer & Apirion, 1981).

RNaseIII cuts in the stems that bracket the mature 16S and 23S rRNA sequences (Bram et al., 1980; Young & Steitz, 1978), and the products are precursor ribonucleoproteins containing RNAs that are slightly longer than the mature molecules. Several attempts have been made to identify conserved sequence elements in primary processing signals (Bram et al., 1980; Gegenheimer & Apirion, 1981; Steege et al., 1987; Daniels et al., 1988; Krinke & Wulff, 1990); these investigations revealed a conserved CUU/GAA box, or a close variation sequence, which exists in most if not all primary substrates. It is also likely that the recognition of an RNA by RNaseIII depends primarily on its secondary and tertiary structure. Since the products of processing by RNaseIII are not the mature rRNA species but precursors that require further processing to complete the maturation process, RNaseIII is not as essential as the other two enzymes, RNaseP and RNaseE (Gegenheimer & Apirion, 1981). It has been demonstrated that enzymes dealing with the precursor, produced after the RNaseIII cleavage, can recognize larger precursors and process them to the normal mature 16S rRNA (Gegenheimer et al., 1977; Sprague & Steitz, 1975). Although in the RNaseIII mutant the processing can still go on, the large size of the substrate molecules can affect the efficiency of their processing by following enzymes; because the enzymes, like RNaseP and RNaseE, find it difficult to access to their processing sites, or allosteric interference obscures the recognition sites. The enzymes that catalyse the processing of pre-16S rRNA and

pre-23S rRNA are ill understood. Purified 16.3S rRNA has a mature 3'-end terminus (Dahlberg et al., 1978) and 66 extra 5' nucleotides, indicating that separate enzymes form the two termini of 16S rRNA.

RNaseE introduces two cleavages, one before and one after the 5S rRNA, which leave a precursor of 5S rRNA (Roy et al., 1983). The catalytic site for RNaseE occurs in an extension of the double-stranded stem which is formed by base pairing of the 5'- and 3'-ends of the 5S rRNA sequence and is known as a molecular stalk (Feunteun et al., 1972).

RNaseP is involved in the cleavages at the 5'-ends of all the tRNA molecules in the nascent rRNA transcript (Altman, 1975). The enzyme generates the mature 5'-terminus of tRNA. Although the maturation of the 3'-end of tRNA is not well understood, it is believed to combine more than two enzymes including both endonucleases and exonucleases (Deutscher, 1984; Watson & Apirion, 1981).

CHAPTER TWO

Materials and Methods

2.1) Bacterial strains and vectors

Mycobacterium smegmatis NCTC 8159 (National Collection of Type Cultures, Public Health Laboratory Services), and *M.smegmatis* mc²155 (obtained from W. Jacobs, Albert Einstein College of Medicine, New York; see Snapper et al., 1990) were held in the laboratory and subcultured on slopes of Löwenstein-Jensen medium (Mackie and McCartney, 1949) at monthly intervals and stored at room temperature. Cloning of mycobacterial DNA was carried out using *E.coli* strains DH5 α F' genotype: F, Q80dlacZ MI5 recA1 endA1 gyrA96 thi-1 hsdR17(rk-,mk+) supE44 relA1 decR(lacZYAargf) U169.

Plasmid vector pUC19 (Norrandar et al., 1983) was used to clone mycobacterial DNA in *E.coli* strains DH5 α F'. The pBluescript (pKS) vector was used for cloning mycobacterial DNA to be sequenced and transformed into *E.coli* strain DH5 α F'. The shuttle vector pEJ106 (E.Davis, unpublished results) was used to clone mycobacterial DNA into *E.coli* and subsequently into *M.smegmatis* mc²155.

2.2) Media

M.smegmatis was grown overnight at 37°C in 5 ml of Lemco broth (Clarke and Meadow, 1959) containing 0.05% Tween 80. This culture was used to inoculate 200 ml Lemco broth medium containing 0.05% Tween 80. Large cultures were incubated at

37°C with rotary shaking at 150 rpm for 40-48h (Model G25 Shaking Incubator, New Brunswick Scientific Company). *M.smegmatis* for electroporation was grown in modified Dubos liquid medium (Dubos and Davis, 1946) kept in a rotary shaking incubator at 37°C and 150 rpm for 24h. After electroporation *M.smegmatis* was plated on Middlebrook 7H10 agar (Difco) containing 10% (v/v) Dubos medium albumin supplement (Difco) and 10-15µg/ml kanamycin. Purity of cultures was checked using acid-fast staining (Cruickshank, 1965).

E.coli was grown in Luria-Bertani (LB) and 2 x TY medium, with antibiotics added as required at the appropriate concentration (Sambrook et al., 1989). Cultures were incubated in a rotary shaking incubator at 37°C and 200-300 rpm for 18-24h. The composition of all media used is given in appendix A.

2.3) DNA preparation

2.3.1) Large scale DNA preparation

Bacteria carrying the recombinant clone, were grown overnight at 37°C, with agitation, in LB medium containing 50µg/ml of the antibiotic selecting for plasmid resistance, (either Ampicillin or Tetracycline). After incubation, 1 litre of the culture was pelleted by centrifugation at 4000 rpm, 4°C for 30 minutes in a Sorvall GS3 rotor, and resuspended in 40 ml of 1 x glucomix (50 mM glucose, 25 mM Tris HCl pH 8.0, 10 mM EDTA pH 8.0). 80 ml of lysis solution (0.2 M NaOH, 1% SDS) was added, followed by 40 ml of 5 M

potassium acetate (KOAc), pH4.8. After mixing, precipitated bacterial components (chromosomal DNA, proteins) were pelleted at 4000 rpm for 30 minutes at 20°C using a Sorvall 6x50 ml rotor and the supernatant was filtered through a gauze. 0.6 volumes of isopropanol were added and after gentle shaking for 15 minutes at room temperature, the recombinant plasmid DNA was precipitated by centrifugation at 4 Krpm for 30 minutes at 15°C. The pellet was washed once with 70% ethanol, dried and resuspended in TE buffer (10 mM Tris HCl pH 7.5, 1 mM EDTA) to a final weight of 9 grams. Ten grams of CsCl and 1 ml of ethidium bromide (stock solution: 5 mg/ml) were added, the solution was mixed well, spun at 2000 rpm for 10 minutes at 20°C and the supernatant was transferred to a Quick-Seal polyallomer centrifuge tube (Beckman). Preparations were spun at 55000 rpm for 36 hours at 20°C in a Beckman Ti70.1 rotor and the band representing the supercoiled form of plasmid DNA was collected. Two volumes of distilled H₂O were added to the DNA solution and the nucleic acid was precipitated by the addition of two volumes of ethanol, washed twice in 70% ethanol, dried and resuspended in TE buffer. RNase A was added to a final concentration of 0.1 mg/ml, and the mixture was incubated at 37°C for one hour. After incubation, the mixture was extracted twice with an equal volume of phenol/CHCl₃ (Sambrook et al., 1989) and ethanol precipitated. The nucleic acid concentration was determined by spectrophotometric determination at 260 nm and 280 nm (normal ratio 260/280=1.8) and its quality and purity were

tested by restriction enzyme digestion on an agarose gel.

2.3.2) Small scale DNA preparation

Individual bacterial colonies were picked from agar plates and grown overnight at 37°C in 10 ml of LB medium containing 50 µg/ml of the relevant selecting antibiotic. The bacteria were pelleted by centrifugation at 2000 rpm for 15 minutes at 4°C, and resuspended in 200 µl of 1 x glucomix. 400 µl of lysis solution was added, followed by the addition of 200 µl of 5 M KOAc, pH 4.8. The precipitate was collected by centrifugation at 15000 rpm for 15 minutes at room temperature and the plasmid DNA was precipitated from the supernatant by the addition of 0.6 volume of isopropanol. The nucleic acid was pelleted by centrifugation at 15000 rpm for 30 minutes at 20°C. The pellet was resuspended in 400 µl of TE buffer, digested with 0.1 mg/ml RNase A at 37°C for 30 minutes, phenol extracted once and ethanol precipitated after the addition of salt. The final pellet was resuspended in distilled H₂O and analyzed by restriction enzymes.

2.3.3) Preparation of *M.smegmatis* total genomic DNA

Genomic DNA was prepared from *M.smegmatis* using a method adapted from that described by Katoch and Cox (1986). *M.smegmatis* was grown to late log phase (40 hours in Lemco broth) and 0.1 volumes of sterile 2 M glycine were then added. The cultures were then incubated overnight, harvested by centrifugation and washed twice in SET buffer (0.3 M sucrose, 50 mM Tris HCl pH 8, 10 mM EDTA). The cells were resuspended in 10 ml SET per g wet weight, lysozyme (Sigma) and lipase type VII (Sigma) were added to give a final

concentration of 2 mg/ml for each enzyme and the suspension was incubated at 37°C for 2h. The cells were recovered by centrifugation and resuspended thoroughly in the minimum volume of SET buffer possible.

Lysis was effected by the addition of 10 ml per g wet weight guanidinium chloride solution (GCB; 6 M guanidinium chloride, 0.1% Tween 80, 10 mM EDTA, 1 mM 2-mercaptoethanol) and resuspending the cells with the aid of a 20 ml syringe. The cells were passed through the syringe gently (to avoid shearing the DNA) about 15 times at room temperature until lysis was complete, at which point the suspension became clear and viscous. Cellular debris was removed by centrifugation at 25000 x g for 10 minutes and the supernatant was extracted with an equal volume of chloroform in glass Corex tubes. The two phases were separated by centrifugation at 7000 x g for 10 minutes and the top phase containing the DNA was removed. DNA was precipitated by carefully overlaying the extracted solution with 2 volumes 100% ethanol and mixing the two phases slowly using a rotary shaker (Orbital Mixer, Denley). This caused the DNA to form a large single clump which was removed using a Gilson P1000 pipette (blue tip), washed 5 times in 70% (v/v) ethanol and blotted dry with Whatman 3 MM paper. The DNA was then dissolved in 10 ml proteinase K buffer (10 mM Tris HCl pH 7.5, 10 mM EDTA, 1% SDS, 0.5 mg/ml proteinase K) to destroy any remaining protein and left to dissolve at 37°C with shaking. When the DNA was completely dissolved the solution was cooled to 4°C to precipitate the SDS which was removed by

centrifugation.

To obtain high purity DNA, the DNA in solution was treated by the methods given in Sambrook et al. (1989). The solution was extracted twice with an equal volume of neutral phenol-chloroform to remove residual protein, and once with chloroform alone to remove residual phenol. RNA was destroyed by incubation with 20000 units RNase T1 (Life Technologies) overnight at 37°C in TE buffer. The RNase was removed by phenol-chloroform and chloroform extractions and the DNA was then ethanol precipitated (using 0.1 volume 3 M sodium acetate and 2 volumes ethanol) and finally dissolved in 100 µl TE buffer. The purity and molecular size range was estimated by agarose gel electrophoresis and the concentration was determined by measurement of the absorbance at 260 nm. Using this method 2.7 mg of high purity DNA was obtained from 7.2 g wet weight *M.smegmatis*.

2.4) RNA preparation

2.4.1) Preparation of *E.coli* total RNA

10 ml Log-phase cells were collected by centrifugation and resuspended in 15 mM Tris HCl, 0.45 M sucrose, 8 mM EDTA, 1% (w/v) SDS and 100 µg/ml proteinase K. After incubation for 15 minutes on ice, the protoplasts were harvested by centrifugation and resuspended in 10 mM Tris HCl, 10 mM NaCl, 1 mM sodium citrate and 1.5% (w/v) SDS, mixed gently and incubated at 37°C for 5 minutes. SDS, protein and DNA were precipitated by adding 250 µl saturated NaCl. After centrifuging, the supernatant was mixed with 2.5

volume of ice-cold ethanol. The precipitated RNA was washed with 70% ethanol, and then redissolved in DEPC (diethyl pyrocarbonate)-treated water. The RNA was stored at -70°C until use.

2.4.2) Preparation of *M.smegmatis* total RNA

M.smegmatis NCTC 8159 or mc²155 was harvested by centrifugation, resuspended in lysis buffer (6M guanidinium chloride, 0.1% v/v Tween 80, 1mM 2-mercaptoethanol, 10mM-EDTA, pH7.0) and passed through a French pressure cell at 18,000 lbf/in² at -10 °C. Insoluble debris was removed by centrifugation at 5000g for 10 minutes and the supernatant was deproteinized by repeated extraction with chloroform/iso-amylalcohol (24:1,v/v). RNA was precipitated by the dropwise addition of a 0.5 volume of ethanol.

2.5) Transcription start site(s) determined by primer extension

Primer extension is used to map and quantitate the 5'-termini of RNA and to detect precursors and processing of mRNA or pre-rRNA. The test RNA is hybridized with an excess of a single-stranded DNA primer (a synthetic oligonucleotide or a restriction fragment) radiolabelled at its 5'-terminus. Reverse transcriptase is then used to extend this primer to produce cDNA complementary to the RNA template. The length of the resulting end-labelled cDNA, as measured by electrophoresis through a polyacrylamide gel under denaturing conditions, reflects the distance between the end-labelled nucleotide of the primer and the 5'-terminus of

the RNA. The yield of cDNA is approximately proportional to the concentration of the target sequences in the mRNA or pre-rRNA preparation.

2.5.1) Preparation of labelled oligonucleotides

Oligodeoxyribonucleotides **P-5** and **KK4.1** (see p65 and p74) were labelled at their 5'-termini as described in Maniatis (Maniatis et al., 1982). Briefly, 10 pmol of oligodeoxyribonucleotide was end labelled in 50 mM Tris HCl (pH 7.6), 10 mM MgCl₂, 5 mM dithiothreitol, 100 μM spermidine, 100 μM EDTA, 50 mCi [γ ³²P] ATP and 15-20 units of T4 polynucleotide kinase in 20 μl final volume. The labelled oligodeoxyribonucleotide was separated from unincorporated radioactive label by gel-filtration through Sephadex G-50 column.

2.5.2) Primer extension

The oligonucleotide primer was end-labelled with ³²P by means of T4 polynucleotide kinase, then gel purified and dissolved in double-distilled H₂O. The ³²P labelled primer (10 ng) was added to 7 μl of water containing 20-30 μg of total RNA. The mixture was denatured at 97°C for 10 minutes and 2 μl of 2 x annealing buffer {50 mM Pipes (piperazine-N,N'-bis[2-ethane-sulfonic acid]) pH 6.4, 2M NaCl} was added. The nucleic acids were annealed at 52°C for 30 minutes, then at 42°C for 20 minutes. 90 μl of reverse transcriptase buffer (60 mM NaCl, 10 mM Tris HCl pH 8.3, 10 mM DTT, 8 mM MgCl₂, 1 mM [dCTP, dGTP, dATP and dTTP] and 50 μg/ml actinomycin D) containing 100 unit/ml Avian Myeloblastosis Virus (AMV)

reverse transcriptase (Pharmacia) was then added to each reaction. The reaction mixture was incubated at 42°C for 1 hour, extracted twice with phenol/CHCl₃, ethanol precipitated and washed twice with 70% ethanol. The extension products were separated on a 12% polyacrylamide-urea gel and visualized by autoradiography.

2.6) Polymerase chain reaction (PCR)

The polymerase chain reaction (PCR) is used to amplify a segment of DNA that lies between regions of a known (or predicted) sequence. Two oligonucleotides are used as primers for a series of synthetic reactions that are catalyzed by a DNA polymerase. The oligonucleotides typically are complementary to sequences that lie on opposite strands of the target DNA, and flank the segment of DNA to be amplified. The template DNA is first denatured by heating in the presence of a large molar excess of each of the two oligonucleotides and the four dNTPs. This mixture is then cooled to a temperature suitable for annealing of the primers to the DNA. Next the primers are extended by DNA polymerase. The cycle of denaturation, annealing and extension/synthesis is repeated many times, each cycle doubling the amount of DNA carried over from the previous cycle. Recent protocols (Saiki et al., 1988), unlike earlier ones, make use of a thermostable polymerase isolated from the extreme thermophile *Thermus aquaticus* (Chien et al., 1976). The thermostability of the enzyme after exposure to the high denaturation temperature, make it unnecessary to

replenish the reaction mixture with DNA polymerase after each successive cycle.

2.6.1) Conventional PCR

1-100 ng of bacterial DNA was subjected to PCR in a total volume of 50 μ l, with 1 unit of *Taq* polymerase (Advanced Biotechnologies LTD), 50 mM KCl, 10 mM Tris HCl pH 8.3, 1.5 mM MgCl₂, 0.01% gelatin, 100 pmol of each of two appropriate primers and 200 μ M of each dNTP (dATP, dCTP, dGTP, dTTP). The 50 μ l mixture was covered by 50 μ l of light mineral oil. The relevant gene fragment coding for Spacer-1 was synthesized using primer combination 1 (JCA3 5'-GCC AAG GCA TCC ACC ATG C-3', based on the *M.tuberculosis* 16S rRNA sequence at positions 2452-2470 (Kempell et al., 1992); and RAC5.1 5'-ATT GAC GGG GGC CCG CAC AAG CG-3', complementary to the 5'-end of the *M.tuberculosis* 23S rRNA gene at positions 1532-1554). The gene fragment coding for Leader was synthesized using primer combination 2 (primer L1 5'-GGG TTG CCC CGA AGC G-3', made from Box B sequence; and primer P₄ 5'-CAC TGC TGC CTC CCG TAG GAG T-3', complementary to the *M.tuberculosis* 16S rRNA gene sequence at positions 960-978). Amplification was achieved using 36 cycles. The reaction mixture was heated to 94°C for 1 minute kept at 58°C for 1 minute then heated to 72°C for 2.5 minutes. This cycle was repeated 35 more times. Finally, the solution was heated to 94°C for 15 seconds, then kept at 58°C for 1 minute and at 72°C for 5 minutes.

2.6.2) Inverse PCR

Genomic DNA of *M.smegmatis* NCTC 8159 was digested with

*Bam*HI for 2 hours at 37°C. Circularization was performed in the same tube with T4 DNA ligase in a diluted DNA concentration that favours the formation of monomeric circles. Ligase activity was destroyed by increasing the temperature to 68°C for 15 minutes. After phenol extraction and ethanol precipitation, the digested and circularized DNA was resuspended in 10 µl water. Two primers (RAC 7.1: 5'-GTG CCA GCA GCC GCG GTA ATA CG-3', made from the *M.tuberculosis* 16S rRNA gene sequence at positions 1126-1148; and P₄: 5'-CAC TGC TGC CTC CCG TAG GAG T -3', complementary to the *M.tuberculosis* 16S rRNA gene sequences at positions 960-978) were used, with their 3'-end facing the unknown region. The other conditions were the same as normal PCR.

2.7) DNA cloning

2.7.1) Restriction enzyme digests and alkaline phosphatase treatment

Restriction enzyme digests of DNA were carried out using a ten-fold excess of the required enzyme in the buffer specified by the manufacturer (Life Technologies) for 1 hour at 37°C. Calf intestinal alkaline phosphatase (CIAP) was used to catalyse the removal of 5' phosphate groups from linearised vector DNA substrate. Up to 50 pmole (as a rough guide 1 pmole of 5'-ends of linear DNA of 4.36 kbp = 1.6 µg) was included in the following reaction: -50 mM Tris HCl pH 8.0, 0.1 mM EDTA pH 8.0, 0.5 units CIAP. The reactions were incubated at 37°C for 30 minutes in a final volume of 20-100 µl. The CIAP was inactivated by heating at 65°C for 30-45

minutes prior to extraction with phenol and precipitation with ethanol. This treatment prevents the religation of the vector unless an insert has been ligated into it.

2.7.2) Blunt-ending of DNA fragments

Where blunt-ended fragments were required for cloning, the following protocol was followed; 1 μ g DNA fragment was added to 3 μ l ligation buffer (Boehringer Mannheim) and 1 μ l (1 unit) Klenow enzyme (Boehringer Mannheim) in a total volume of 30 μ l, and incubated at 37°C for 8 minutes. dNTP mix (2 μ l of 0.2 mM of each of the four nucleotides) was then added and the reaction left at 37°C for a further 10 minutes. The DNA was then purified by phenol-chloroform extraction and ethanol precipitation.

2.7.3) Purification of DNA fragments by agarose gel electrophoresis

Restriction endonuclease-digested genomic and vector DNA was usually purified prior to cloning using agarose gel electrophoresis (Sambrook et al., 1989). One of two methods was used to extract the DNA from the agarose after electrophoresis:

(a) After electrophoresis in a TAE (40 mM Tris-acetate pH 7.7, 1 mM EDTA) buffer system, DNA was extracted from the agarose gel using the glassmilk binding method. The manufacturer's protocol (BIO 101) was followed.

The DNA band was sliced out of the agarose gel, 3 volumes of sodium iodide solution were added and the mix incubated at 50°C until the agarose was completely dissolved. 5 μ l (plus 1 μ l extra for each μ g of DNA over 5 μ g) of the

glassmilk solution was added, and the contents mixed by inverting the tube several times. The mixture was incubated on ice for 5 minutes. The glassmilk was recovered by spinning in a microcentrifuge for 1 minute, and washed three times in 500 μ l of the ethanol wash provided. A volume of TE equal to that of the glassmilk was added and the precipitate resuspended. After incubation at 50°C for 5 minutes the solution was spun and the TE containing the DNA was recovered. This step was repeated once and the supernatant pooled. DNA thus recovered was suitable for further manipulation.

(b) The alternative method, which involved running the DNA onto a strip of Whatman 3 MM paper, was an adaption of the method described by Girvitz et al. (1980). The 3 MM paper (Whatman) was wetted in electrophoresis buffer and laid onto a strip of dialysis membrane. A slit was cut in the gel in front of the DNA band to be purified and the 3 MM paper/dialysis membrane was inserted into the slit, with the paper facing the DNA. A second strip of paper plus membrane was inserted behind the band to be purified, so that the dialysis membrane face prevented any DNA situated behind the band of interest from moving. The gel was then subjected to further electrophoresis until the DNA required had migrated completely onto the paper.

The strip was then removed, the dialysis membrane discarded, and the 3 MM paper placed inside a 0.5 ml microcentrifuge tube with a small hole pierced in its base. This tube was then placed into a 1.5 ml microcentrifuge tube

and spun at 13000 rpm for 5 minutes. The DNA solution was collected and the extraction was repeated twice by washing the paper with a volume of 1 x TAE equal to that spun off the paper. The DNA washes were then pooled, extracted with neutral phenolchloroform and ethanol precipitated.

2.7.4) Making a vector for PCR products cloning

Bluescript plasmid was digested with *EcoRV* and incubated with *Taq* polymerase (1 unit/ μ g plasmid/20 μ l vol.) using standard buffer conditions (50 mM KCl, 10 mM Tris HCl pH 8.3, 1.5 mM MgCl₂ and 200 μ g/ml BSA) in the presence of 2 mM dTTP for 2 hours at 70°C. The addition of a single thymidine at the 3'-end of each restricted vector facilitated the ligation with PCR products which are found to have an overhanging adenosine residue at the 3'-end (Clark, 1988). After phenol extraction and ethanol precipitation, the vector was ready for cloning.

2.7.5) Ligations

Ligations were set up in a total volume of 10 μ l containing 25 ng of vector DNA, 50-200 ng of mycobacterial DNA, 1 unit of T4 DNA ligase (Boehringer Mannheim) and 1 μ l of 10 x commercial ligation buffer (Boehringer Mannheim). These were left at room temperature overnight before use.

2.8) Transformation of *E.coli*

Competent *E.coli*, strain DH5 α F', cells were prepared using an adaptation of the method described by Cohen et al. (1972). A single colony of *E.coli* maintained on minimal medium was inoculated into 5 ml LB broth and grown overnight

at 37°C. A 1/100 dilution of culture was used to inoculate 50 ml LB broth (in a 250 ml conical flask) and grown at 37°C with rapid rotary shaking (250 rpm) until the optical density of the culture at 550 nm was between 0.3-0.4. The cells were chilled on ice for 10 minutes prior to harvesting by centrifugation (4°C, 2000 x g for 10 minutes). The supernatant was decanted and the cells were resuspended in 10 ml of ice cold 0.1 M CaCl₂. The cells were then sedimented by centrifugation and resuspended in 4 ml of 0.1 M CaCl₂. The suspension was stored on ice for at least 2 hours and an aliquot of 200 µl taken for each transformation.

Between 5 and 100 ng of DNA was added to each aliquot of cells and the mixture kept on ice for a further 30 minutes. The cells were heat-shocked at 42°C for 90 s and then 800 µl of LB broth was added to the mixture, which was incubated at 37°C for 1 h. Transformed cells were plated onto selective media in dilutions which gave 100-300 colonies per plate. When pUC19 or pBluescript SK was the cloning vector, plates contained 50 µg/ml ampicillin, 0.2 M isopropyl-β-D-thiogalactopyranoside (IPTG) and 200 µg/ml 4-bromo-3-chloro-2-indolyl-β-galactoside (X-Gal). This combination facilitates blue-white screening: colonies with recombinant plasmids are white and parental vector transformants blue. When the shuttle vectors were used, selection was based on resistance to 50 µg/ml ampicillin or kanamycin.

2.9) Electroporation of mycobacteria

Electroporation of mycobacteria was carried out using

an adaptation of the method described by Snapper et al. (1988). One ml of an overnight culture of *M. smegmatis* mc²155 was used to inoculate 100 ml of modified Dubos medium and this was shaken at 150 rpm at 37°C for 24 h. The cells were harvested by centrifugation (2000 x *g* for 10 minutes), washed in 10% (v/v) glycerol and resuspended in 10 ml 10% (v/v) glycerol.

Plasmid DNA, up to 1 µg, was added to 0.8 ml of the cell suspension and placed in an electroporation cuvette (0.2 cm electrode gap, BioRad Laboratories) on ice for 10 minutes. The cells were then given a single pulse of 25 µF at 6250 V/cm with resistance set at '∞' (Gene Pulser Apparatus, BioRad Laboratories) and replaced on ice for 10 minutes. An equal volume of Lemco broth was added and the cells incubated at 37°C for 2 h. Cells were pelleted in a microcentrifuge at 13000 rpm for 5 minutes, resuspended in 100 µl Lemco broth and plated on 7H10 agar plates containing 15 µg/µl kanamycin.

2.10) Agarose gel electrophoresis

Electrophoresis in agarose gels was used to separate, measure the size of and purify DNA fragments, according to the methods described in Sambrook et al. (1989). Several different sized gel plates were used. Two sizes were made at the NIMR, namely 6x12 cm and 10x15 cm. A commercial mini-gel (6x8 cm; Horizon 58, BRL) and midi-gel (12.8x15 cm; Midi I, Northumbria Biologicals Ltd) were also used with suitable comb sizes. Various concentrations of agarose were used for

the gel matrix, depending on the size of the DNA fragments to be resolved. The gel thickness was kept to 3-4 mm. Ethidium bromide was included in the gel itself at a final concentration of 0.25 $\mu\text{g/ml}$ in order to visualise the DNA in UV light. Prior to electrophoresis 1/5 volumes of loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol FF, 30% v/v glycerol) were added to each sample.

Electrophoresis was carried out in one of two buffer systems. When DNA fragments were to be recovered using Geneclean, a Tris-acetate buffer was used (TAE 1x solution - 40 mM Tris-acetate pH 7.7, 1 mM EDTA). For all other purposes a TBE buffer was used (TBE 1x- 89 mM Tris-borate, 2.5 mM EDTA, pH 8.0). After electrophoresis, gels were placed on an ultraviolet transilluminator (360 nm; UV Instruments Ltd). The ethidium bromide-labelled DNA bands were visualised and photographed using Polaroid type 667 film.

Lambda DNA digested with *Hind*III was most commonly used as size markers for linear DNA fragments. The sizes of the fragments are 23.7, 9.5, 6.7, 4.3, 2.25, 1.95 and 0.56 kbp. A wider size range in the range of 0.8-5.24 kbp is generated by double digestion of lambda DNA with *Eco*RI and *Hind*III. The fragments generated are 5.24, 5.0, 4.25, 3.51, 1.99, 1.88, 1.58, 1.35, 0.94 and 0.84 kbps. There is a linear relationship between the logarithm of DNA mobility and agarose gel concentration. The sizes of restriction fragments were estimated by running DNA fragments of known size on the same gel as the restriction fragments. The size

was estimated from the plot of mobility (in mm distance migrated from the well) versus the M.W. of DNA fragments of known size on logarithmic graph paper.

2.11) DNA sequencing

2.11.1) Preparation of clones for sequencing

Initially two restriction fragments were prepared from the 1.8 kbp fragment of the construct KEK6 (see **Fig.3.1**); a 678 bp *Pst*I-*Bam*HI fragment and a 1.122 kbp *Bam*HI-*Pst*I fragment (see **Fig.3.2**). Both were cloned into pUC19 vectors (Boehringer Mannheim), which were digested with appropriate enzymes to allow sequences to be determined on both strands.

All the PCR products were cloned into pBluescript vector (Boehringer Mannheim), which was digested with *Eco*RV enzyme and labelled at the 3'-end with a single thymidine (see section 2.7.4).

The above recombinant DNAs were used to transform 200 μ l aliquots of competent *E.coli* DH5 α F', prepared by the CaCl₂ method (section 2.8). Small scale DNA preparation was prepared as described in section 2.3.2.

2.11.2) Preparation and sequencing of double-strand DNA

Double stranded DNA was denatured in the presence of 0.2 M NaOH and 0.2 mM EDTA for 30 minutes. After adding 0.1 volume of 3 M sodium acetate (NaOAc) pH 5.2, the DNA was precipitated by ethanol.

Sequencing reactions were carried out using the Sequenase 2.0 kit (US Biochemical Corporation). Primers were annealed to denatured plasmid DNA as follows: 7 μ l DNA, 2 μ l

5x reaction mix and 1 μ l appropriate primer were incubated at 60°C for 15 minutes and cooled at room temperature for 5 minutes before use. Labelling: the labelling mix was diluted 5 fold with sterile distilled water (SDW) and stored frozen at -20°C until use. The SequenaseTM enzyme was diluted 1:8 in enzyme dilution buffer immediately prior to use. Once the template-primer reaction had reached a temperature below 30°C, the reaction was transferred onto ice. The following reagents were then added to the tubes: 1 μ l of DTT, 2 μ l of diluted labelling mix, 0.5 μ l of [α -³⁵S] dATP, and 2 μ l of diluted SequenaseTM (added last). After mixing thoroughly, the mixture was incubated at room temperature for 2-5 minutes. Termination: For each reaction, four tubes were labelled G, A, T and C. To each tube was added 2.5 μ l of the appropriate dideoxy NTP termination mix (eg. ddATP termination mix to tube A). Before adding the labelling mix, the tubes were prewarmed to 37°C while the labelling reaction was taking place. When complete, 3.5 μ l of the labelling mix was transferred to each tube and incubation was continued for a further 3-5 minutes. 4 μ l of stop solution was then added to each tube, the contents mixed thoroughly and stored on ice ready to load onto a sequencing gel (occasionally they were stored at -20°C for up to 1 week). When the gel was ready for loading, the samples were heated to 75-80°C for 2 minutes and loaded immediately using 2-3 μ l per lane.

2.11.3) Preparation and electrophoresis of sequencing gels

Products of the sequencing reactions were separated by polyacrylamide gel electrophoresis under denaturing

conditions using a 50 cm x 20 cm gel apparatus (Cambridge Electrophoresis Equipment) and also using 60 cm x 21.6 cm gel plates for the Base Runner apparatus (International Biotechnologies Inc.). The glass plates were thoroughly cleaned with a dilute solution of Decon 90 (Decon Laboratories Ltd), rinsed several times with distilled water followed by ethanol and left to dry. The notched plate was siliconised by spreading about 10 ml dimethyldichlorosilane solution (2% solution in 1,1,1-trichloroethane, BDH) using a Kleenex paper towel, rinsed twice with distilled water and then with ethanol. The gel was then assembled using Whatman 3MM paper as spacers and held in place with clips.

Acrylamide was obtained as a 40% stock solution (38% acrylamide, 2% bis-acrylamide, Amresco, Camlab) and TBE buffer was prepared as a 10x stock and diluted to 1x as required (1x TBE - 89 mM Tris-borate, 2.5 mM EDTA, pH 8.0). Top and bottom acrylamide solutions were freshly prepared for buffer gradient gels. These gels allow a larger number of bases to be read from a single gel by reducing the migration speed of DNA at the base of the gel, thereby decreasing band spacing. The top solution consisted of 6% acrylamide, 8 M urea, and 0.5x TBE, with the bottom solution containing 6% acrylamide, 8 M urea and 5x TBE.

For the smaller gels (45 cm length plates) 50 ml of top solution was primed with 100 μ l of 25% (w/v) ammonium persulphate (APS) and 100 μ l N,N,N',N'-tetramethylethylenediamine (TEMED). For the longer 60 cm plates, 70 ml of top solution was primed with 120 μ l of 25%

(w/v) APS and 120 μ l TEMED. For both sets 10 ml of bottom solution was primed with 15 μ l of 25% (w/v) APS and 15 μ l TEMED. For both size plates 6 ml of top solution and 7 ml of bottom solution were drawn into a 25 ml pipette with a few air bubbles to aid mixing and poured down one side of the gel. The remaining top solution was then poured to fill the gel mould and a comb fitted in place. Gels were left to polymerize for 45 minutes before the gel apparatus was assembled.

Once the gels were in contact with the running buffer (1x TBE) the combs were removed and the gels were pre-run at constant power for 15-30 minutes (37 W for 45 cm plates and 45 W for 60 cm plates). Sequencing reaction mixtures (4 μ l) were carefully loaded using a Gilson P20 micropipette and 'duckbill' tips (Gilson, Anachem) and then gels run at the same constant power as for the pre-run. Samples were loaded in the order T,C,G,A from left to right. For sequences from 0-200 bases the gel was run until the bottom dye reached the base of the plate (approximately 4 h); for sequences from 200-400 bases the longer 60 cm gels were run at reduced power (35 W) for 12 h.

After electrophoresis the glass plates were separated, with the gel remaining stuck to the unsiliconised plate, and the acrylamide gel containing DNA fragments was "fixed" in 10% (v/v) methanol, 10% (v/v) glacial acetic acid for 15 minutes. The gel was then removed from the backing plate by overlaying it with a sheet of 3MM paper and carefully lifting. The gel was overlaid with 'clingfilm' and dried

under vacuum on a gel drier (BioRad laboratories) at 80°C for 45 minutes. The dried gel was exposed in direct contact with Fuji X-ray film for 24-48 hours before the autoradiograph was developed (Fuji Automatic X-ray Film Developer). Autoradiographs were read using a gel reader (IBI Gel Reader, International Biotechnologies Inc.) and sequences fed directly into the MacVector computer program.

CHAPTER THREE

Promoter Activity of the *rrn* Operons of *M.leprae* and *M.tuberculosis*

3.1) Introduction

It has been demonstrated that *M.leprae* and *M.tuberculosis*, like other slow-growing mycobacteria, contain only a single set of rRNA genes (Sela et al., 1989 and Bercovier et al., 1986). It is shown in this thesis that there are no difference in the single *rrn* operon possessed by both species that are sufficient to explain the dramatic difference in growth rate of these two pathogens (12 days and 24h respectively). There is a broad correlation between the rate of growth of a bacterium and the number of its ribosomes (Bremer & Dennis, 1987). In turn, the number of ribosomes is determined by the production of rRNA by *rrn* operons. Nevertheless, in addition to the number of *rrn* operons, the rate at which mature rRNAs are produced depends also on the features of their promoters which are thought to be responsible for growth rate and stringent control (Ota et al., 1979; Miura et al., 1981), and the efficiency with which the operons are transcribed and processed. These factors are more important to the slow-growing mycobacteria, which have only one *rrn* operon at their disposal.

This chapter describes (1) the determination of the promoter regions of the *rrn* operons of *M.leprae* and

M.tuberculosis, (2) investigation into their ability to function as promoters in *E.coli*, and (3) identification of the transcription start site in *E.coli*.

3.2) Results

3.2.1) Cloning and sequencing of the 5' region of the *M.leprae* *rrn* operon

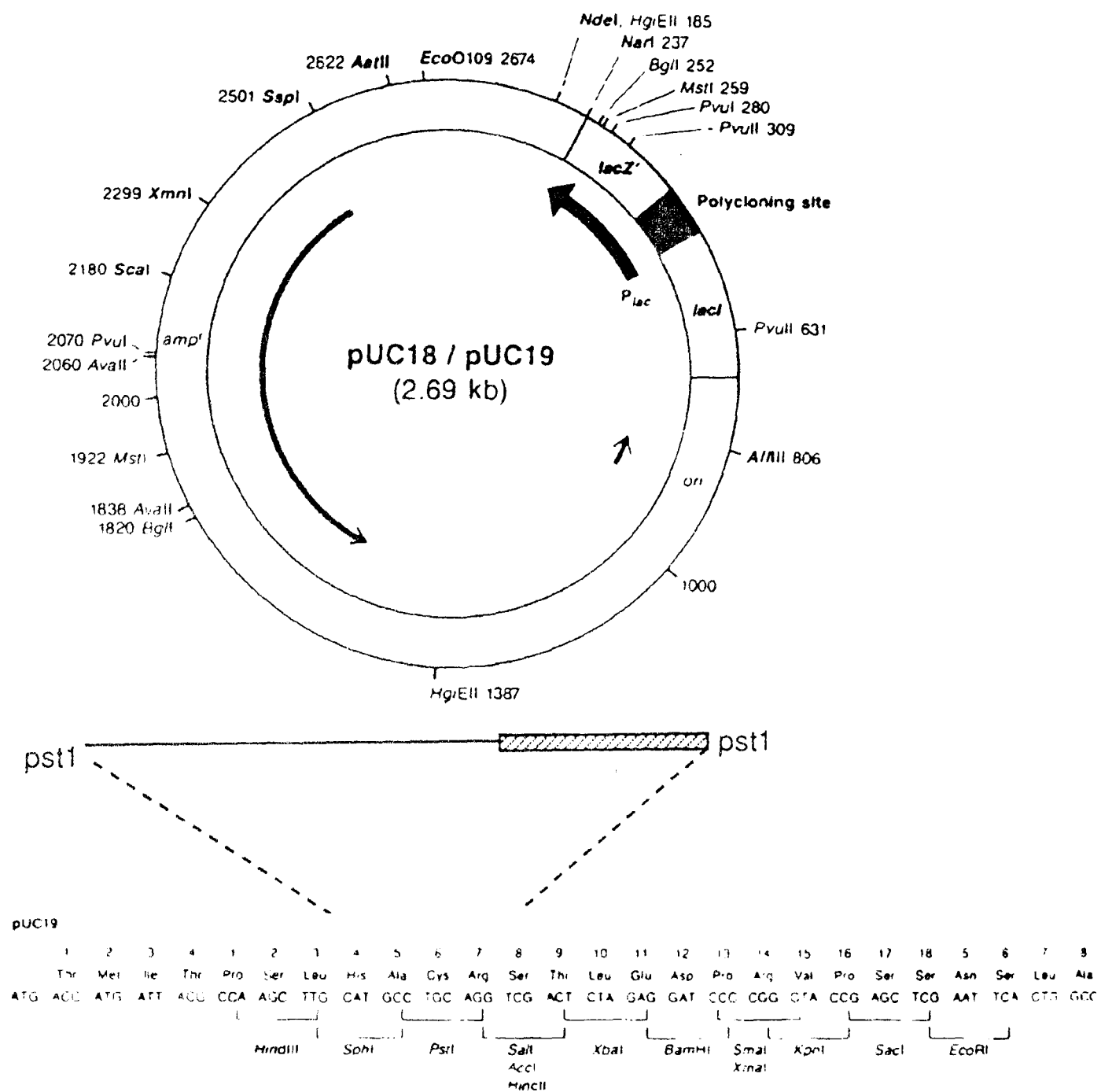
To study the function of the promoter and the leader regions of the *rrn* operon of *M.leprae*, the whole 5'-flanking region was cloned and sequenced. The previously cloned plasmid KEK6 (Fig.3.1) has a 1.8 kbp *Pst*I insert which contains approx.0.6 kbp of the 16S rRNA gene and 1.2 kbp of the 5'-upstream flanking region. The sequencing strategy used is shown in Fig.3.2. First the 1.8 kbp fragment was digested with *Bam*HI enzyme and then the two fragments, KEK6.1 and KEK6.2, with sizes of 685 bp and 1115 bp respectively, were cloned into pBluescript II SK vector. The whole sequence upstream of the 5'-end of 16S rRNA gene was established and is shown in Fig.3.3. The base content of the 5'-flanking region was as follows: 21.9% A, 24.6% T, 24.3% C and 29.2% G. The molar G+C content, 53.5%, was consistent with the G+C content of 57.7% already determined for the same region of *M.tuberculosis* (Kempell et al., 1992). Putative -35 and -10 boxes were identified, at nucleotides 902-907 and 923-928 respectively, with 15 nucleotides between them.

3.2.2) Investigation of promoter activity in *E.coli*

Fragments containing the whole promoter of the *M.leprae*

Fig.3.1 KEK6 plasmid construct containing the 5'- flanking region of *M.leprae* 16S rRNA gene. The *M.leprae* 1.8 kbp *Pst*I fragment was cloned into the pUC 19 vector at the *Pst*I site.

Fig 3.1



200 400 600 800 1000 1200 1400 1600 1800

P **B** **S** **P**
 ↓ ↓ ↓ 16S rRNA gene ↓ **A**

: P₁ : P₂ P₄ :
 : → → ← : → ← :
 : ← ← : ← :
 : : : :
 : : : :
 : : : : **B**

1115 bp

↓
 pBluescript II SK

↓
 pBluescript II SK

(A) shows the construct KEK6 containing the 1.8 kbp cloning fragment. The solid portion represents the 16S rRNA coding region, showing restriction sites for P- *Pst*I, B- *Bam*HI, S- *Sau*3A. The arrows indicate the positions of primers used in sequencing. (B) shows the two fragments subcloned into pBluescript II SK. The arrows indicate the positions of the primers used in sequencing.

```

1   ATGACAACCA GCGCGGCGGC ACGCTGCAAT GGGTTAGGCC GGATGGTGCC
51  GACGTGTTTCG AGGCGAAAAC GATGGTTGGG ACGCGGATGT CGGCGCAGAT
101 CTCCTCATAC ACGTCGAAGA TGGTGTCGAC TTTGGCATCA CATCACTATA
151 GTAAATGGCA GGTTAGTGGC CAACCCAACG GAAAGTAGGC GCCGACGATT
      P-1
201 TCAGCCAGCT GTTCGCGGGT GTAGTTTAGC GTGTCCGTAA GAACCCGCCG
251 GCGCACCAAT GGCGCGGGTC GTGGCTGTGT CCAAATAAGG ATTAAAGACA
301 AGATAAAGAC AACGTGATGG TGCCAACCCA TGGTGAGCTG TACACCCAGC
351 AGATATTGAT CCCACCTAG CGCAGCATTT TGTCGCCCTC GCCGGGACGG
401 CGTCGGTGGA CAACTGCGGA TTAGAGATTC GTAGTTGCGC AGCAGGATCA
451 TTAGATCGCC ACGCGGTCGA TCAACCAGCG GTCGAAACAC CGGGTCGAAA
501 ACCATCTCGG AACATGTGGT CAGACCCACA CGGTTGAATC GGGTGCAATC
551 GACGCGCAGC ATTGCCGGAT AGTCGCTGGG TTCGATGGCA GCGTCTAACA
601 ATGGAAACAC GCGCCAGCTT CCTCGGCCGT GCCGCTGAGT TCAAACAGAG
651 CTATCACGGC CGTACTTAAC GTCCTTAGGA TCCGGCGTGT CTGTGTTCAA
701 CCCGGCACGC CGTGCCTCGT GCGGGTTGAA ATAGGCCTTG TTCCCAGAAT
751 TGCGGTGATC ACCAACGGCT CGCCCGGCGC AATCGCAATG TCGTCTGGGG
801 CAGTCCGGAC CGCAGTAGTA GGTCATAACT GTTGAGGTAC ACGCCGGCCG
851 CTTGCCTTGC GGCGGTCCTC ATGACGAATC TCCTCGACGA CTTCATCGAC
      P-2
      -35box      Sau3A  -10box      *      Box B
901 ATTGTGTACG GTGACAGATC TGTATTAATC TGGCTGGGTT GCCCCGAAGC
      PL1
951 GGGGGAAGTA AGCTTGAAGT GTTGTTTGAG AACTCAATAG TGTGTTTGGT
1001 TTTGTTGTTG TTGATTTTTT GACTACATCT AGCATTCCTC GTGTGTGTAG
1051 GTGTAGTTTA TTATGTTATT TATAGATGCC AGTTTTGGTG TCTTGTGAGG
1101 TATCTCTAGA AATTGAAAAT TTCGTCTAGT TATTGATGGA GTT

```

Figure 3.3: The sequence of the 5'-flanking region of the *M. leprae* 16S rRNA gene. The promoter -10 and -35 elements are underlined, as are the two primers used to identify the cloning artifact upstream of the -10 box. *, represents the transcription start site. The letters in italics indicate a *Sau3A* restriction site. Primer **PL1** made from the putative Box B element is underlined. Sequence starts from 5' to 3'.

and *M.tuberculosis* *rrn* operons were each subcloned in both orientations into the shuttle vector pEJ106 (Fig.3.4) at a *Bam*HI restriction site. This vector is suitable for studying promoter activity because it contains a CAT (chloramphenicol acetyltransferase) gene preceded by a multiple cloning site and can replicate both in *E.coli* and *M.smegmatis* (Davis,EO; unpublished results). The inserted promoter thus controls CAT gene expression, and expression of the CAT gene under the control of a promoter of correct orientation allows bacteria to grow in the presence of chloramphenicol. The clones with the *M.leprae* promoter in correct and wrong orientations are named as pJY101 and pJY102 respectively; the clones with the *M.tuberculosis* promoter in correct and wrong orientations are named pJY104 and pJY103 respectively. Transformed *E.coli* were selected on LB medium containing 34 µg/ml chloramphenicol, and subsequently grown on medium containing 170 µg/ml chloramphenicol. Only *E.coli* transformed with the plasmids pJY101 and pJY104, containing the promoter element in the correct orientation to allow the expression of the CAT gene, formed colonies in the chloramphenicol medium (Table 3.1). The resistance to the high concentration of chloramphenicol produced may reflect the high copy number of pJY101 and pJY104 in *E.coli* or, alternatively, high expression of the CAT gene by the inserted promoter. It is considered that the latter factor might play an important role, because (i) analysis of the sequences from both *M.leprae* and *M.tuberculosis* *rrn* operons revealed a

Fig.3.4 Construction of pEJ106 derivatives, namely pJY 101, pJY 102, pJY 103 and pJY 104. Their inserts are illustrated in Appendix B. pEJ 106 (Davis, EO unpublished result) is a 8.9 kbp plasmid, which contains a CAT gene lacking its own promoter, preceded by a multiple cloning site. Plasmid pEJ 106 bears two selective markers, ampicillin (Amp) and kanamycin resistance (Kan). E, *EcoRI*; Sm, *SmaI*; B, *BamHI*; Sa, *SalI*; Ps, *PstI*; H, *HindIII*. *BamHI* is the unique restriction site in pEJ 106 plasmid. The pEJ 106 derived plasmids were used to transform *E.coli*, and the resultant cells were grown in LB medium with different concentrations of chloramphenicol.

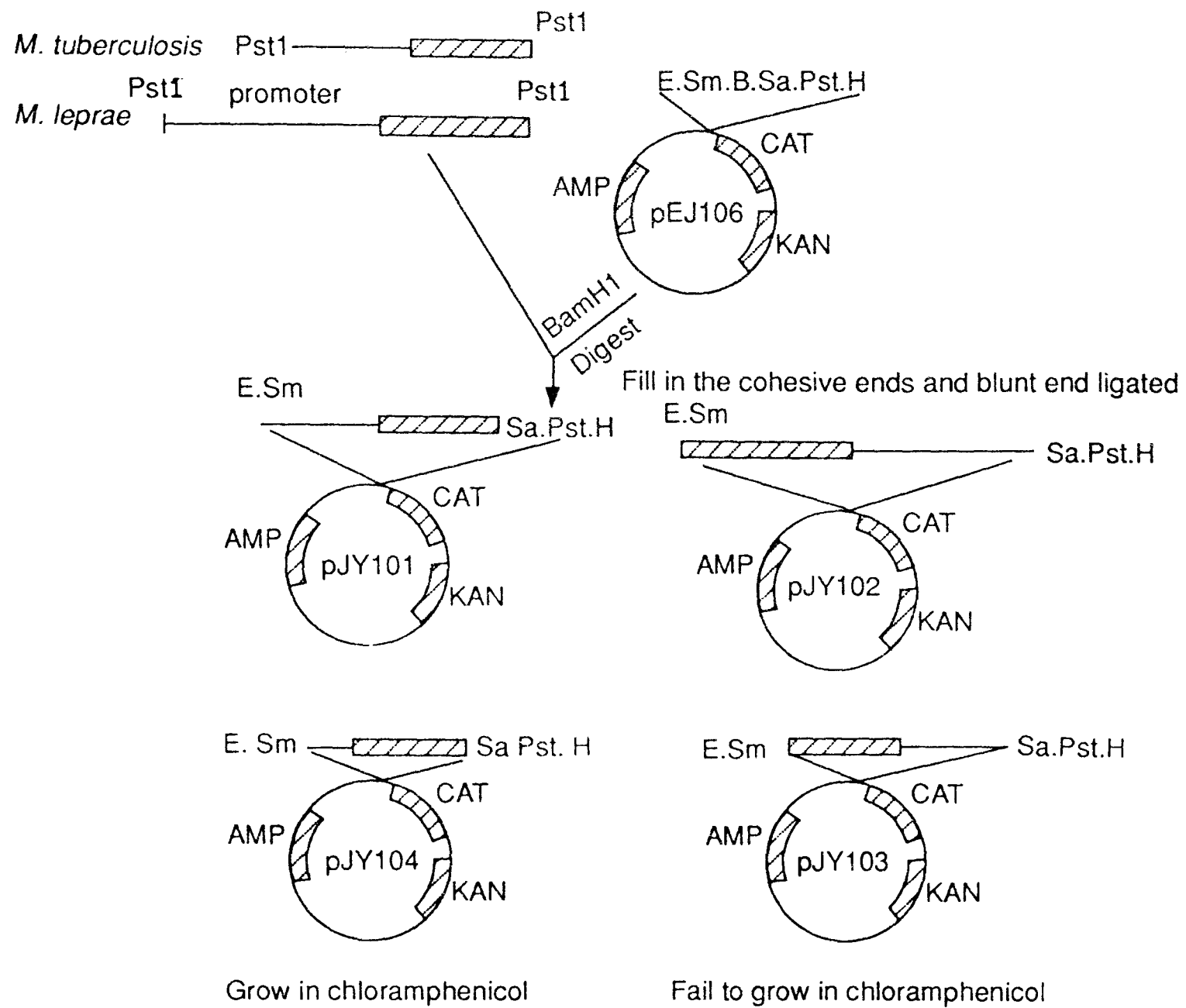


Table 3.1 The promoters of *M.tuberculosis* and *M.leprae* *rrn* operons function in *E.coli*

Plasmids used for transformation	Ampicillin 50 µg/ml	Chloramphenicol	
		34 µg/ml	170 µg/ml
pJY 101	+	+	+
pJY 102	+	-	-
pJY 103	+	-	-
pJY 104	+	+	+

promoter-like element which is close to the canonical -10 and -35 regions found in many bacteria (Harley & Reynolds, 1987) and (ii) the sequences of the two hexamers, which are separated from each other by 15 and 18 nucleotides in *M.leprae* and *M.tuberculosis* respectively, are similar to the promoter elements in *E.coli*. It is assumed that these sequences are recognized by the *E.coli* transcription system with the result that the downstream reporter CAT gene is expressed in high yield. Analysis of the DNA sequences upstream from the -35 box of both species revealed no potential open reading frame(ORF). Analysis of the reverse-complementary strand did not show any substantial ORFs.

3.2.3) Recognition of a cloning artefact

3.2.3.1) The method used to prepare a *M.leprae* DNA cosmid library

By screening a *M.leprae* cosmid library provided by Ms Jenny Thompson, London School of Hygiene and Tropical Medicine, a cosmid recombinant KEK5 was identified which contained the entire *M.leprae* *rrn* operon. The library was prepared as follows: A partial *Sau3A* digested *M.leprae* DNA was cloned into the *Bam*HI restriction site of the cosmid vector Lorist4 (Cross & Little, 1986). Recombinant DNA was packaged *in vitro* before transfecting *E.coli* strain 1046 grown in the presence of kanamycin (Sigma) 25 µg/ml. KEK6 plasmid was subcloned from KEK5.

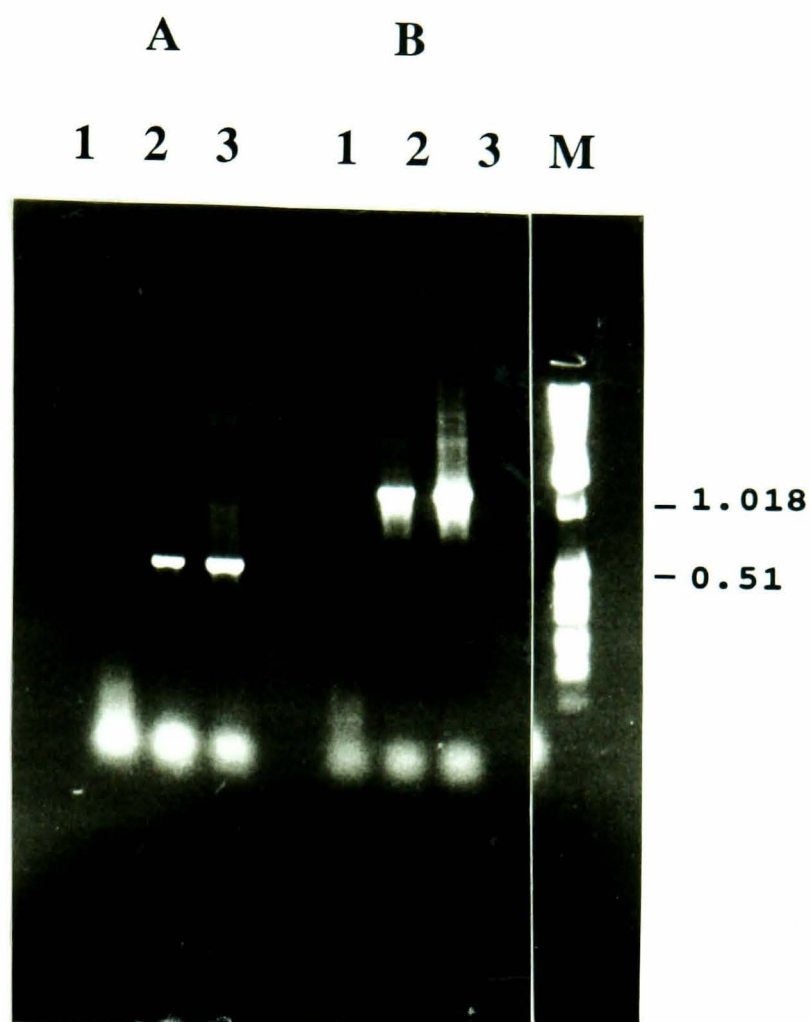
3.2.3.2) Identification of the sequence artefact upstream of the -10 box

After sequencing approximately 1.2 kbp of the 5'-flanking region of the *M.leprae* *rrn* operon, a *Sau*3A restriction site was identified (**Fig.3.3**) at the nucleotide position 917-920. Since the library used was constructed by partial digestion by the enzyme *Sau*3A, it was possible that the *Sau*3A site was either genuine, or an artefact due to ligation of non-related fragments.

PCR amplification was employed to test the latter possibility. Two primers (shown in **Fig.3.3**) made from the sequences upstream of the *Sau*3A restriction site were used, each combined with the primer P4 made from the 16S rRNA gene. *M.leprae* genomic DNA, KEK5 cosmid and KEK6 plasmid were used as substrates. If the cloned sequence upstream of the *Sau*3A site was correct and present in the *M.leprae* genome, then by PCR a band of amplification would be expected with each of the substrates. Fig.3.5 shows that amplification did not occur with *M.leprae* genomic DNA. This result indicates that the recombinant plasmid sequence arises from a cloning artefact upstream of the -10 box, starting from a *Sau*3A restriction site, and is due to chance ligation of an unrelated fragment.

After the cloning and characterization of the promoter of the *M.leprae* *rrn* operon had been carried out, other workers (Sela & Clark-Curtiss, 1991) reported similar results for the expression of the promoter in *E.coli*. Their sequence data support our conclusion that the sequence upstream of the *Sau*3A restriction site in the plasmid clone KEK6 is due to a cloning artefact. The genuine -35 box is

Fig.3.5 Identifying cloning artefact(s) by PCR



Primer **P-1** 5' TTA GTG GCC AAC CCA ACG GA 3'

P-2 5' CTT CAT CGA CAT TGT GTA CGG 3'

P-4 5' CAC TGC TGC CTC CCG TAG GAG T 3'

Figure 3.5 Polymerase chain reaction experiment to identify a cloning artefact. The primers used in identification of the cloning artefact are indicated and their positions are indicated by arrows in **Figure 3.2**. Primer combination **P-2/P-4** was used in reactions of group (**A**), and primer combination **P-1/P-4** was used in reactions of group (**B**). **1**, *M.leprae* genomic DNA was used; **2**, KEK5 cosmid was used; **3**, KEK6 plasmid was used; and **M**, Lambda markers, size given in kbp.

16 nucleotides upstream of the -10 box with a sequence TTGACT, which has 2 nucleotide differences from the false one in KEK6. The genuine promoter also functions strongly in *E.coli*. It is likely that the promoter of the mycobacterial *rrn* operon, and the near-fit to it that occurs slightly further downstream, might provide all the activating function necessary for expression of the operon.

3.2.4) Expression of *M.tuberculosis* *rrn* in *E.coli*

The promoter of the *M.tuberculosis* *rrn* operon functions in *E.coli*, which means its transcriptional signal can be recognized by the *E.coli* system. Since the region around the putative promoter possesses sequences that have high G+C contents, it was of interest to see transcription structure. Also the identified transcription start site would help to locate the position of the genuine promoter.

3.2.4.1) 5'-end of the *rrn* operon of *M.tuberculosis*

The transcription start point of the *rrn* operon is marked by the 5'-end of pre-rRNA. An attempt was made to identify this site by the primer extension method (see Materials and Methods) using *M.tuberculosis* rRNA (25 µg/assay) as the substrate. No product was detected.

An alternative strategy was then used, namely to express the *rrn* operon in *E.coli* by means of the promoterless CAT plasmid pEJ106 (see section 3.2.2) and to identify the 5'-end of the transcript controlled by the *M.tuberculosis* *rrn* operon promoter introduced into the plasmid. Accordingly, the PIE1 plasmid having the 1.2 kbp

*Pst*I fragment which contains the promoter region (Fig.3.6) and part of the 16S rRNA gene of the mycobacterial *rrn* operon, was recloned into the pEJ106 plasmid as described in Methods. The recombinant plasmid, named pJY104, was used to transform *E.coli* and transformants were identified by their resistance to chloramphenicol. The RNA fraction was isolated and used as a substrate in the primer extension assay. The primer P-5 was made from the leader region of *M.tuberculosis rrn* operon and used in the primer extension experiment. Its sequence and position are shown in Fig.3.6. The results presented in Fig.3.7 indicate a leader region of 191 nucleotides. The transcription start site is found near the 5'-end of the Box B element identified previously (Kempsell et al., 1992). This site also corresponds to the 5'-end of the primer L1 which was used to amplify the leader region of *rrn* sequences (see next chapter).

3.2.4.2) Comparison with the 5'-end of the *M.leprae rrn* operon

The primer extension experiment (Fig.3.7) shows that the *rrn* operon of *M.tuberculosis*, expressed in *E.coli*, has a single promoter which is situated close to the Box B element. A similar result was reported recently for the promoter activity of the *M.leprae rrn* operon (Sela & Clark-Curtiss, 1991). There are clear homologies between the promoters of *rrn* operons of *M.tuberculosis* and *M.leprae* (Table 3.2). The transcription start sites of the two operons, both located immediately downstream of the -10 box, indicate that the putative *rrn* operon promoter elements

```

1      CTGCAGCGAT GCTATCGCTT TGGCGTCGAT CGCCGACGGC ACATCGATGA
51     TCACGGAGAA CGTGTTTCGAG GCGCGGTTCG CGCGCGGTTC CGCTTCGTTG
101    AAGAGATGAT CCGGCTCGGT GCATACGCTC GGACCGACGG GCACCACGCC
151    GTGGTGCGGG GCCTCCCGCA GCTGTCGAGC GCTCCGGTGT GGTGTTTCGGA
201    CATCCGTGCC GGGGCCGGCT TGGTGCTGGC GGGGCTCGTT GCCGACGGCG
251    ACACCGAGGT CCACGATGTA TTCCACATCG ATCGCGGATA TCCGTTGTTC
301    GTGGAGAACC TGGTGAGTCT CGGTGCCGAG ATCGAACGGG TATGCTGTTA
                                     -35box
351    GGCGACGGTC ACCTATGGAT ATCTATGGAT GACCGAACCT GGTCTTGACT
                                     -10box      *   Box B
401    CCATTGCCGG ATTTGTATTA GACTGGCAGG GTGCCCCGA AGCGGGCGCA
                                     PL1
451    AACAAAGCAAG CGTGTTGTTT GAGAACTCAA TAGTGTGTTT GGTGGTTTCA
                                     P-5
501    CATTTTTGTT GTTATTTTTG GCCATGCTCT TGATGCCCCG TTGTCGGGGG
551    GCGTGGCCGT TTGTTTTGTC AGGATATTTC TAAATACCTT TGGCTCCCTT
601    TTCCAAAGGG AGTGTTTGGG T

```

Figure 3.6: The sequence of the 5'-flanking region of *M.tuberculosis* 16S rRNA gene. The promoter elements, -35 and -10 boxes, are indicated. *, represents the transcription start site. The primer (P-5) used for primer extension experiments is underlined, as is the primer PL1 that is made from the putative Box B element. The nucleotide sequence is from 5'-end to 3'-end.

Fig. 3.7 Transcription start site of *M.tuberculosis* *rrn* operon in *E.coli*

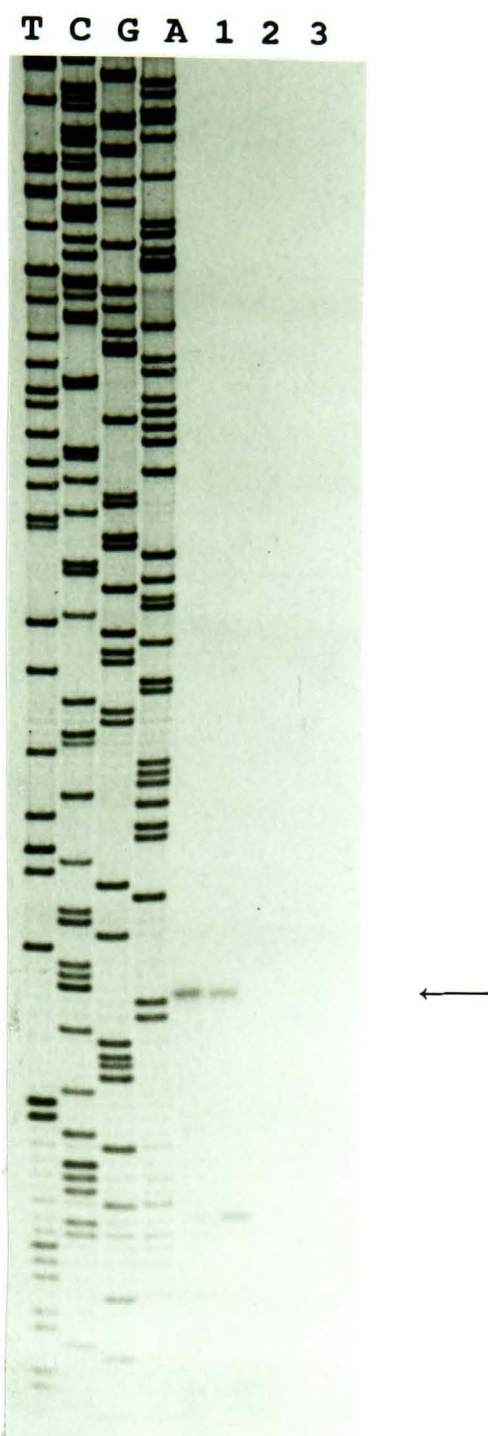


Figure 3.7 Primer extension analysis of the *M.tuberculosis* promoter. Lanes **T**, **C**, **G** and **A** represent sequencing reactions of the pJY104 insert using primer **P-5** (see Fig.3.6). Lanes 1-3 represent primer **P-5** extension reactions, done with RNA (12 μ g) purified from *E.coli* cells harbouring plasmid pJY104, RNA (24 μ g) purified from *E.coli* cells harbouring plasmid pJY104 and pEJ106 (parent vector of pJY104), respectively. The arrow indicates primer extension products of interest. The faster moving bands in lane 1 and 2 correspond to a partial product terminating close to the end of the Box B sequence.

identified in both *M.leprae* and *M.tuberculosis* are genuine. There are also homologies between the mycobacterial promoters and the second promoter (P2) of the *rrnO* operon of *B.subtilis* (Table 3.2). However, the transcription start point in the *M.tuberculosis rrn* operon differs from that of the *M.leprae rrn* operon, in that in *M.leprae* the G residue is the expected start site. The observed start site is within one nucleotide, a difference which is within experimental error.

3.2.5) Expression of the *M.tuberculosis rrn* operon in *M.smegmatis*

Although only one transcription start point was detected when both *M.leprae* and *M.tuberculosis rrn* operon promoters were expressed in *E.coli*, it is still possible that another promoter might exist and function in mycobacteria. Tandem promoters precede the 16S rRNA gene of some other species, and it is possible that they also occur in *M.leprae* and *M.tuberculosis*, but that only one is recognized by *E.coli*. This possibility was tested by introducing pJY104 constructs into the mycobacterial system.

After transformation of *M.smegmatis* mc²155 with the pJY104 construct by electroporation, the transformants were identified by their resistance to kanamycin. 200 ml Lemco Broth culture was inoculated with the transformants and shaken until the OD reached 0.5. The RNA fraction was isolated, and used as substrate in the primer extension

Table 3.2. Comparison of *rrn* operon promoter sequences of *M.leprae*, *M.tuberculosis*, *B.subtilis* and *E.coli*.

Species	Gene	-35 box	-10 box		Transcription start site*
<i>E.coli</i>	<i>rrn</i> G P ₁	TTGTCA	TATAAT	GCGCCACC	ACTGACAC
	<i>rrn</i> G P ₂	TTGACT	TATTAT	GCACACG	GCCGCGCC
<i>B.subtilis</i>	<i>rrn</i> O P ₁	TTTACA	TATAAT	CATTTCT	GTTGTCT
	<i>rrn</i> O P ₂	TTGACC	TACTAT	TAAGTA	GTCGCTT
<i>M.leprae</i>	<i>rrn</i>	TTGACT	TAATCT	GGCTGG	GTTGCCCC
<i>M.tuberculosis</i>	<i>rrn</i>	TTGACT	TAGACT	GGCAGG	(G) TTGCCCC

*, the first nucleotide represents the transcription start site.

References for sequences are as follows: *rrnG*, Lindahl & Zengel, 1986; *rrnP*, Ogasawara et al., 1983; *M.leprae rrn*, Sela & Clark-Curtiss, 1991.

assay (for detail see Materials and Methods section). The primer extension experiment is shown in Fig.3.8. An extra band was detected in RNA isolated from mc²155 transformed with pJY104 when compared with the control RNA isolated from mc²155 transformed with pJY104's parent vector pEJ106. The extra bands detected in lane 1 and lane 2 are at the position where the transcription start site was detected in *E.coli* (see lane 5). Thus the result supports the conclusion that only one promoter exists in the *M.tuberculosis rrn* operon.

3.3) Discussion

The work in this chapter has helped to establish the following points: (1) the identified actual promoters of *rrn* operons in both *M.leprae* and *M.tuberculosis* function in an *E.coli* system; (2) the difference in growth rate between *M.tuberculosis* and *M.leprae* is unlikely to be due to the different features of the promoters of their *rrn* operons; (3) the pre-rRNA molecules are rapidly processed even in slow-growing mycobacteria; (4) the -10 box of the *rrn* operon of slow-growing mycobacteria is important for the expression of the mycobacterial *rrn* operon promoter in *E.coli*.

The slow growth rates of the pathogenic mycobacteria hampers study of the control of the expression of their genes. It has been found that mycobacterial genes are not generally expressed in *E.coli* and *B.subtilis*, possibly because their regulatory signals can not be recognized by

Fig. 3.8 Transcription start site of *M.tuberculosis* *rrn* operon in *M.smegmatis*



Figure 3.8 Identification of the transcription start site of the *rrn* operon of *M.tuberculosis* expressed in *M.smegmatis* mc²155. Lanes 1-5 represent primer **P-5** extension results. Lane 1, extension product from RNA (12 µg) isolated from mc²155 transformed with plasmid pJY104; lane 2, as for lane 1 but using 24 µg RNA substrate; lane 3, primer **P-5** labelled with ³²P; lane 4, extension products using RNA (20 µg) isolated from mc²155 transformed by pEJ106, parent vector of pJY104; lane 5, done with RNA purified from *E.coli* cells harbouring plasmid pJY104. The arrow indicates the extension products of interest. Other bands are extension products from RNA of mc²155 or partial products terminating before the actual transcription start sites. The **T**, **C**, **G** and **A** are sequencing reactions of the pJY104 insert using the primer **P-5**.

the RNA polymerase system of the heterologous cells. That the identified promoters are actual *M.leprae* and *M.tuberculosis* *rrn* promoters is supported by the observation of promoter activity in *E.coli*, sequence similarity to the consensus promoter, and the position of the transcription start site of the *rrn* operon of both species, which are located 8 bp downstream from the identified -10 box. According to this study, promoters from 16S rRNA genes of both *M.leprae* and *M.tuberculosis* direct transcription in *E.coli*. The possible explanation for the expression of the mycobacterial *rrn* operons in *E.coli* is that from the evolutionary point of view, the transcription system remains universal in living cells and the promoter of the *rrn* operon may be conserved because of the fundamental function of rRNA in protein biosynthesis. This conservation means that the *rrn* promoter elements can form a universal three dimensional structure which can be recognized by both *E.coli* and mycobacterial polymerase. That makes it possible to study the regulation of the expression of either the *M.leprae* *rrn* operon or the *M.tuberculosis* *rrn* operon in a heterologous host, such as *E.coli*. Since the promoter of the *M.tuberculosis* *rrn* operon has been successfully expressed in the fast-growing mycobacterium *M.smegmatis*, the *M.smegmatis* system offers an alternative, promising approach for studying the regulation of rRNA synthesis in slow-growing mycobacteria.

Although *M.leprae* cannot be grown *in vitro*, and *M.tuberculosis* grows very slowly, the *rrn* operons of

M.leprae and *M.tuberculosis* are expressed in *E.coli* from their own promoters, so it is not surprising that a good match with the *E.coli* consensus promoter is found upstream of the 16S rRNA gene of each of these two mycobacteria (see **Table 3.2**). In addition, in both the species of mycobacteria, the transcription start sites for the *rrn* operons are both located at the conserved Box B element, and they differ by only one nucleotide. Thus it is unlikely that the exceptionally slow-growth rate of *M.leprae* arises from a defect in its capacity for rRNA biosynthesis.

Using RNA directly isolated from *M.tuberculosis* I failed to detect pre-rRNA by means of the primer extension method, which suggests that it is not feasible to use the method to study slow-growing mycobacterial pre-rRNA. This negative result suggests that processing rapidly follows transcription so that the ratio pre-rRNA:rRNA is likely to be inversely proportional to the doubling time of the bacterium. In *E.coli* cells, processing is rapid and most rRNA is mature. Only 1-2% of rRNA is present as long precursors (King & Schlessinger, 1983). For slow-growing mycobacteria like *M.tuberculosis* it appears that pre-rRNA is not sufficiently abundant to be detected by the primer extension procedure.

Although the cloning artefact occurred between the -10 box and the -35 box, the whole promoter is still functional and the downstream CAT gene is expressed. There are several possibilities to explain this phenomenon. Firstly, there may be a -35 box-like element in a suitable

place which compensates for the whole function of the *rrn* operon. By examination of the sequence, a TTGTGT element at the nucleotide position 902-907 was found which is 16 nucleotides upstream of the -10 box. The near optimal distance and the two nucleotide differences compared with the -35 box in the *E.coli rrn* operon, make it able to function as a genuine -35 box. Secondly, it can be predicted that in the *M.leprae rrn* operon the -35 box is not as important as the -10 box when it is expressed in an *E.coli* system. It also implies that the upstream sequence has little influence on the *M.leprae rrn* operon promoter activity when it is expressed in *E.coli*.

CHAPTER FOUR

Sequence Analysis of the Leader Region of the rRNA Gene of

Slow-Growing Mycobacteria

4.1) Introduction

The transcription start site of the *rrn* operon of *M.tuberculosis* was identified in Chapter 3. This chapter concerns the leader region, which begins at the transcription start site and extends to the 5'-end of the 16S rRNA gene. The leader sequence contains a number of elements considered to be of primary importance for processing and transcription antitermination (Gourse et al., 1986; Holben et al., 1985; Aksoy et al., 1984). The Box A, Box B and Box C elements in phage λ and *E.coli* *rrn* operons are thought to be responsible for the end-to-end transcription of the long untranslated rRNA genes without premature termination (Li et al., 1984; Berg et al., 1989). Since little is known of the anti-termination mechanism in mycobacteria, it is of interest to sequence the leader region of the mycobacterial *rrn* operon. By identifying sequence homology with *E.coli*, I hoped to establish whether the same mechanism exists in the slow-growing mycobacteria.

In *E.coli*, the leader region of *rrn* operon is found to play a role in determining the secondary structure of pre-rRNA. The secondary structure influences the cleavage of the pre-rRNA by RNases to produce the mature rRNAs.

Establishing the nucleotide sequence for the leader region of slow-growing mycobacteria should help to define processing of mycobacterial pre-RNA.

It is known that sequence similarity in 16S rRNA genes among slow-growing mycobacteria is above 95% (Rogall et al., 1990b). Thus it is difficult to distinguish between closely related species, for example within the *M.tuberculosis* complex. Since the growth rate among slow-growing mycobacteria differs dramatically, it was possible that the leader regions of mycobacterial *rrn* operons would reveal more sequence differences than shown by the 16S rRNA gene.

4.2) Formulation of the hypothesis: the transcription start site is conserved among slow-growing mycobacteria

Analysis of sequences spanning positions around P1 promoter transcription start sites of *rrn* operons in *E.coli* reveals a conserved sequence, the GC rich discriminator, which is thought to be responsible for stringent control (Travers, 1980a & 1984). The results of the experiments described in Chapter 3 suggested that only a single promoter exists for the *rrn* operon of the slow-growing mycobacteria. It is likely, therefore, that this single promoter is subject to stringent control and a similar conserved GC rich sequence in the leader region of slow-growing mycobacteria *rrn* operon was predicted.

Comparison of the leader sequence of the rRNA gene of *M.leprae* with the same region in *M.tuberculosis* (Fig.3.3

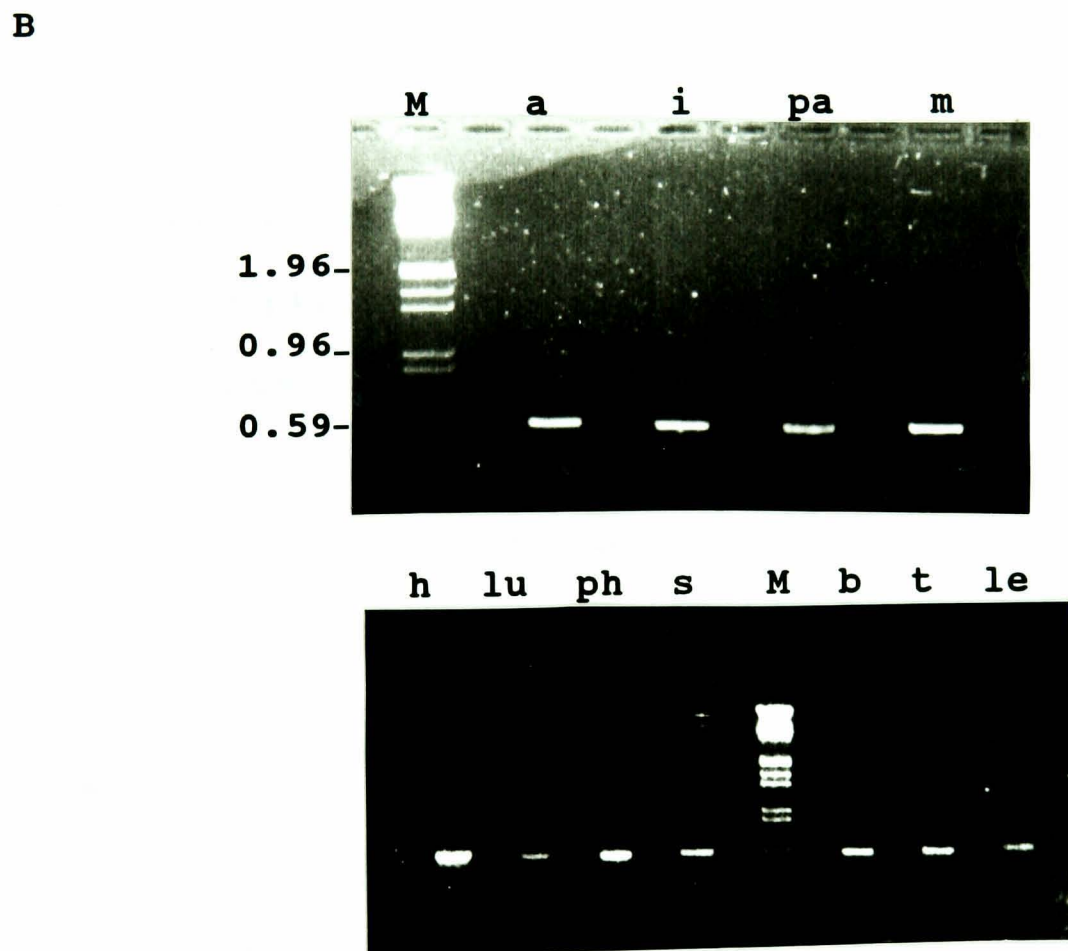
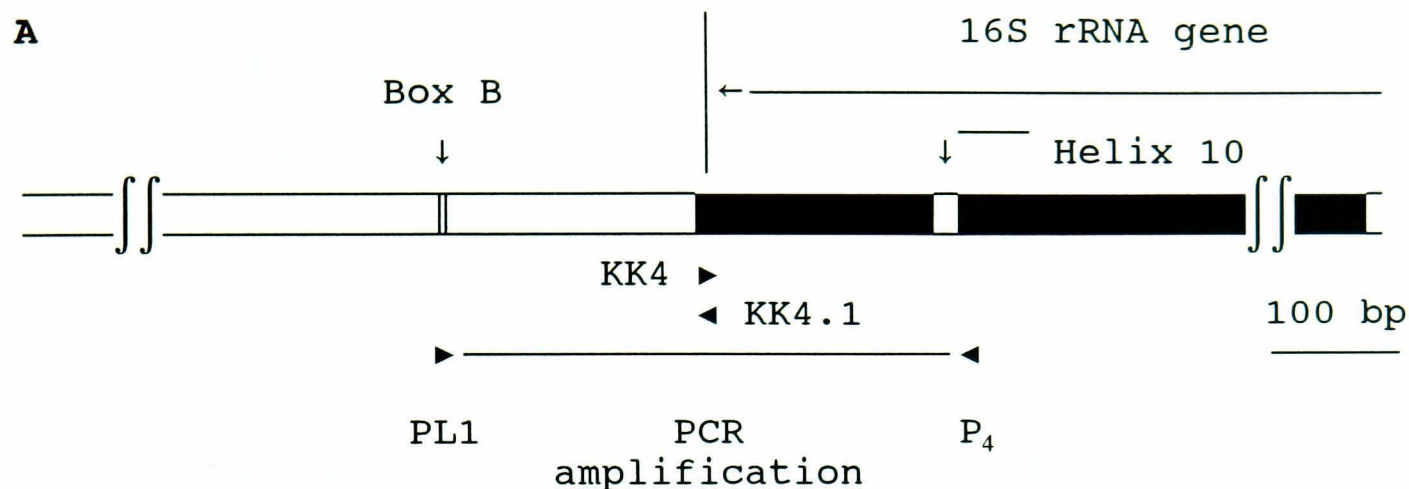
and 3.6), revealed that an identical sequence, similar to the Box B element of the *nutL* gene of bacteriophage lambda (Friedman & Gottesman, 1983), was present in both species. Furthermore, the transcription start sites of the *rrn* operon in both species were located near to the 5'-end of the putative Box B element (Fig.3.3 and 3.6). The possibility that slow-growing mycobacteria have very similar control elements of the *rrn* operon was investigated on the basis of the assumption that if the Box B element is highly conserved among slow-growing mycobacteria, then the leader regions of slow-growers could be amplified by PCR utilizing shared sequences as primers.

4.3) Results

4.3.1) Amplification and cloning of the leader region of *rrn* operons of slow-growing mycobacteria

Based on the above hypothesis, PCR was used to amplify the leader region of the *rrn* operon of slow-growing mycobacteria (see Fig.4.1). Briefly, a primer (primer L1) from Box B, with a sequence 5'-GGGTTGCCCCGAAGCG-3', was synthesized. This primer was used, together with another primer (primer P4) made from conserved region of the 16S rRNA gene with a sequence 5'-CACTGCTGCCTCCCGTAGGAGT-3', in the PCR amplification using genomic DNA of slow-growing mycobacteria as templates (see Materials and Methods). Each of the nine slow-growing mycobacteria tested yielded a

Amplification of the leader regions of mycobacteria



C Primer PL1 see Figure 3.6.
P₄ See Figure 3.5.

Figure 4.1 (A) Polymerase chain reaction strategy to amplify the leader region of the *rrn* operons of slow-growing mycobacteria. (B) The PCR products were run on 1% agarose gel and stained with ethidium bromide. (C) Primers used in the PCR experiment. M- Lambda/*EcoRI*+*HindIII* digested markers, size given in kbp; a- *M.avium*; i- *M.intracellulare*; pa- *M.paratuberculosis*; m- *M.marinum*; h- *M.habana*; lu- *M.lufu*; ph- *M.phlei*; s- *M.smegmatis*, b- *M.bovis*; t-*M.tuberculosis*; le- *M.leprae*.

product of about 580 bp long, which contained about 200 bp nucleotides of the leader sequence (**Fig.4.1**). All the products were directly cloned into the *EcoRV* restriction site of pBluescript vector as described in Materials and Methods, and the recombinant DNAs were used to transform *E.coli* DH5 α F' strain. Transformants were selected on LB medium with the antibiotic ampicillin at 50 μ g/ml.

4.3.2) Sequencing of the leader regions of *rrn* operons of slow-growing mycobacteria

At least three colonies from the same ligation reaction were selected, cultured and plasmid DNA isolated. Sequences of the inserted DNAs were determined as described in the Chapter 2. The above procedure reduces the possible errors caused by PCR amplification (Saiki et al., 1988). Primers used to perform the sequencing were universal, reverse primers of pBluescript and KK4/KK4.1 (see **Fig.4.1**). Between 150 and 300 bp of nucleotide sequence immediately flanking the above primers was obtained. The sequence of the 16S rRNA gene served to confirm the identity of the species studied. The sequence of helix 10 of the V2 region of 16S rRNA gene (see the scheme for secondary structure of Kempsell et al., 1992), which is a useful signature of the sequence (Rogall et al., 1990a), is given in Table 4.1. The length of the leader region of the species studied is presented in Table 4.2. The leader region sequences of nine slow-growing mycobacteria and their 5'-ends of the 16S rRNA gene were established. The leader sequences were aligned by

Table 4.1 The sequences of helix 10 of the V2 regions confirm the species studied*.

	179		202
<i>M.avium</i>	ACCUCAAGAC	GCAU	GUCUUCUGGU
<i>M.paratuberculosis</i>	ACCUCAAGAC	GCAU	GUGUUCUGGU
<i>M.intracellulare</i>	ACCUUUAGGC	GCAU	GUCUUUAGGU
<i>M.lufu</i>	ACCUUAAGAC	GCAU	GUCUUUUGGU
<i>M.simiae</i> (<i>M.habana</i>)	ACCACUUGGC	GCAU	GCCUUGUGGU
<i>M.marinum</i>	ACCACGGGAU	UCAU	GUCCUGUGGU
<i>M.leprae</i>	ACUUCAAGGU	GCAU	GUCUUGUGGU
<i>M.tuberculosis</i>	ACCACGGGAU	GCAU	GUCUUGUGGU
<i>M.smegmatis</i>	CCUGCUGGUC	GCAU	GGCCUGGUAG

*. The nucleotide position and helix designation are based on the scheme of secondary structure of the *M.tuberculosis* 16S rRNA (Kempsell et al., 1992).

M.leprae. See also Sela & Clark-Curtiss., 1991 and Cox et al., 1991.

M.tuberculosis. Kempsell et al., 1992.

The first nucleotide shown corresponds to *E.coli* 16S rRNA position 186.

the computer PILEUP program of GCG V.7 (Genetic Computer Group Inc., Madison, Wisconsin) (Fig.4.2).

4.3.3) Identification of functional elements of the leader regions of slow-growing mycobacteria

4.3.3.1) Elements important to the start of transcription

The alignment of the sequences of the slow-growing mycobacteria studied indicated several places where the sequences are conserved among different species. One place is the 5'-end of the leader region, which contains the whole Box B element. The Box B element is conserved among all the species studied, since an appropriate product was obtained in each case. On the basis of previous work (Cox et al., 1991) it is thought that there are unlikely to be more than four mismatches between primer PL1 and its target. As described in Chapter 3, I have identified the transcription start site for the *M.tuberculosis* *rrn* operon. Sela & Clark-Curtiss (1991), have similarly described the site for *M.leprae*. In both species transcription was found to start within the Box B element. Since Box B can always form a stem-loop structure, it may be functioning as a signal to start transcription. It is likely from the structure and the sequence similarity that transcription of other slow-growing mycobacterial *rrn* operons starts within the Box B element as well.

4.3.3.2) Elements important to antitermination

The elements Box A, Box B and Box C were first

Figure 4.2 The alignment of leader sequences of slow-growing mycobacteria.

Alignment of sequences of the leader regions of the *rrn* operons of slow-growing mycobacteria according to the PILEUP program. *M.av*, *M.avium*; *M.pa*, *M.paratuberculosis*; *M.in*, *M.intracellulare*; *M.lu*, *M.lufu*, *M.si*, *M.simiae*; *M.ma*, *M.marinum*; *M.le*, *M.leprae* and *M.tb*, *M.tuberculosis*. Each nucleotide sequence was obtained as described in Materials and Methods using primers: **PL1** and **P₄** for the amplification step. The first nucleotide corresponds to the 5'-end of primer **PL1**. The *M.leprae* sequence has been published by Liesack et al., 1990; Sela & Clark-Curtiss, 1991. The *M.tuberculosis* sequence has been published previously by Kempsell et al., 1992. The available sequence for the *rrn* operon of *M.bovis* (Suzuki et al., 1988) is the same as the *M.tuberculosis* sequence. *, denotes the consensus site for the start of transcription. •, deletion of nucleotide residue. Nucleotides 3-216 (L1-L214) inclusive comprise the leader region of pre-rRNA which extends from the 5'-end of the start of transcription to nucleotide 1 of the 16S rRNA coding region. Box A, Box B and Box C are putative antitermination elements implicated in transcription of the gene. The stem (invariant) sequence is identical in all eight species studied and may interact with the complementary, c (invariant) sequence of ITS to form basepairs. The small letters are used to indicate nucleotide differences from the equivalent *M.avium* sequences. The nucleotide sequences start from 5' to 3'.

	Primer PL1 ← sequence →		← Stem (invariant) →		
	*	Box B		Box A	50 (L48)
<i>M.av.</i>	GGGTTG	CCCCGAAGCGGG	CGAAA•••••C•AAG	TGTTGTTT	GAGAACT
<i>M.pa.</i>	GGGTTG	CCCCGAAGCGGG	CGAAA•••••C•AAG	TGTTGTTT	GAGAACT
<i>M.in.</i>	GGGTTG	CCCCGAAGCGGG	CGAAA•••••C•AAG	TGTTGTTT	GAGAACT
<i>M.lu.</i>	GGGTTG	CCCCGAAGCGGG	CGAAA•••••C•AAG	TGTTGTTT	GAGAACT
<i>M.si.</i>	GGGTTG	CCCCGAAGCGGG	CGAAAacaagcaa•gAG	TGTTGTTT	GAGAACT
<i>M.ma.</i>	GGGTTG	CCCCGAAGCGGG	tGAgtgcaagtaagAAa	TGTTGTTT	GAGAACT
<i>M.le.</i>	GGGTTG	CCCCGAAGCGGG	gGAA•gtaagcttgAAG	TGTTGTTT	GAGAACT
<i>M.tb.</i>	GGGTTG	CCCCGAAGCGGG	CGgAAacaagcaag•cG	TGTTGTTT	GAGAACT
					100 (L98)
<i>M.av.</i>	CAATA	GTGTGTTTG	GTCTTTTTA•TTTGTGTTG••TTTTTGGCCATA•C		
<i>M.pa.</i>	CAATA	GTGTGTTTG	GTCTTTTTA•T•••TTGTTG••TTTTTGGCCATA•C		
<i>M.in.</i>	CAATA	GTGTGTTTG	GTCTTTTTA•TTTGTGTTGttTTTTTGGCCATA•C		
<i>M.lu.</i>	CAATA	GTGTGTTTG	GTCTTTTTA•TTTGTGTTG•tTTTTTGaCCATA•C		
<i>M.si.</i>	CAATA	GTGTGTTTG	GTgTTTTTg•TTTGTGTTGttTTTTTGaCCATA•C		
<i>M.ma.</i>	CAATA	GTGTGTTTG	GTgTTaTgt•TTTGTGTTG•tTTTTTGaCtAcA•C		
<i>M.le.</i>	CAATA	GTGTGTTTG	GTtTTgT•••••TGTTGTTGatTTTTTGaCtAcAtC		
<i>M.tb.</i>	CAATA	GTGTGTTTG	GTggTTTcAcaTTtTTGTTGttaTTTTtGgCcatgC		
					150 (L148)
<i>M.av.</i>	••CTAGCACTCCCCGTGTGTGGGTATGG•CAA•••••TTTATGATGCCAG				
<i>M.pa.</i>	••CTAGCACTCCCCGTGTGTGGGTATGG•CAA•••••TTTATGATGCCAG				
<i>M.in.</i>	••CTAGCACTCCCCGTGTGTGGGTATGG•CAA•••••TTTATGATGCCAG				
<i>M.lu.</i>	••tTcGCACcCCCCGTGTGTGGGTATGGttgt•••••TTTtTGATGCCAG				
<i>M.si.</i>	tttTAaCACTCCCCGTGTGTaGGTATGGtCg•••••TTTtTGATGCCAG				
<i>M.ma.</i>	cttatcacCcCCgtGTGaGgGaGTAgtcgttt•••••TTTtaGATGCCAG				
<i>M.le.</i>	tagcAttcCTtCgtGTGTGTaGGTgTaGtttAttatgTTatTtATagatG				
<i>M.tb.</i>	tctTgatgCccCgttgtcGgGGGgcgtGgCcgtttgtTTTgTcAgGatAt				
					200 (L198)
<i>M.av.</i>	TT•TT••TGGCGTCTTGTCAGGT••••ATCTCTTGATTTGAAATTCACCT				
<i>M.pa.</i>	TT•TT••TGGCGTCTTGTCAGGT••••ATCTC•TGATTTGAAATTCACCT				
<i>M.in.</i>	TTaTT••TGGtGTCTTGTCAGGT••••ATCTC•TGATTTGAAATTCACCT				
<i>M.lu.</i>	TTaaT••TGGCGTCTTGTCAGGT••••ATCTCT•gaTTtgAATTCACCT				
<i>M.si.</i>	TatTTggTGtCtTtTTGTtAGGTcagatTaTCTctgaTTcgAATcCACCT				
<i>M.ma.</i>	TTgaT••TGGCGTCTTGTCAGGTatctcT••••GATTgcgAATTCACCT				
<i>M.le.</i>	ccagTttTGGtGTCTTGTCAGGTatctcTagaaat•TgaaAAatTTCgtCT				
<i>M.tb.</i>	TTcTaaaTacCtTtggcTCcctTttccAaagggaGtgT•••••				
<i>M.av.</i>	CGCTGCGCGAGGAGAT				
<i>M.pa.</i>	CGCTGCGCGAGGAGAT				
<i>M.in.</i>	CGtTcatCGAGGAGtT				
<i>M.lu.</i>	CGtTctGCGAGGAGtT				
<i>M.si.</i>	CGtTGatCGAGGAGAT				
<i>M.ma.</i>	CGtT•atCGAGGgGtT				
<i>M.le.</i>	actTattgatGGAGtT				
<i>M.tb.</i>	•••••				

16S rRNA gene

identified in bacteriophage λ . They were subsequently found in the *rrn* operons of *E.coli*, although in the order Box B, Box A and Box C in the leader region and Box B and Box A (but no Box C) in the spacer region (see, for example, Berg et al., 1989). The three motifs (Box B, Box A and Box C) are found within approximately 70 bp immediately following the *E.coli rrnG* P2 promoter (Li et al., 1984). These motifs have their homologues within 65 bp downstream from the putative start of transcription of the *rrn* operon of the slow-growing mycobacteria studied (**Fig.4.2**).

Box B is a region of hyphenated dyad symmetry with the potential for the RNA-like strand to form a hairpin loop. The sequence 7-19 (5'-CCCCGAAGCGGGC-3'; **Fig.4.2**) is very similar to the Box B element (5'-GCCCTGAAGAAGGGC-3') of the *nutL* gene of bacteriophage λ (Friedman & Gottesman, 1983).

The Box A motif has an even more conserved sequence which is 5'-(C/T)GCTCTT(T)A-3' in bacteriophage λ (Friedman & Gottesman, 1983), 5'-TGCTCTTTA-3' in the *rrnG* operon of *E.coli* (Li et al., 1984) and 5'-AGTTCTTTG-3' in the *rrnO* operon of *B.subtilis* (Ogasawara et al., 1983). Homologous sequences are found in positions 36-44 (5'-TGTTGTTTG-3') of **Fig.4.2**.

The Box C element has the consensus sequence 5'-GG(T/C)GT(G/A)(T/C)G-3' in bacteriophage λ (Friedman & Gottesman, 1983) and 5'-TGTGTGGG-3' in the *rrnG* operon of *E.coli* (Li et al., 1984); the slow-growing mycobacteria homologue appears to be 5'-GTGTGTTTG-3' (positions 56-64, **Fig.4.2**).

4.3.3.3) Elements important to pre-rRNA processing

The stem region (**Fig.4.2**) has a high percentage of homology with its *B.subtilis* counterpart. The *B.subtilis* sequence (Ogasawara et al., 1983) is 5'-GTTCTTTGAAAA↓CT↓AAA-3' (the arrows indicate RNase III cleavage sites) compared with the slow-growing mycobacteria sequence 5'-GTTGTTTGAGAA*CT*CAA-3' (the asterisks indicate possible RNase III cleavage sites). In the *rrn* operons of *B.subtilis* (Stewart & Bott, 1983), *Mycoplasma* (Taschke & Herrman, 1986) and *Streptomyces* (Pernodet et al., 1989), the Box A motif is located within the stem structure close to the RNase III processing site, as found for slow-growing mycobacteria (see **Fig.4.2**).

4.3.3.4) The sequence is more variable in the leader region

One of the striking features of the *rrn* operons of slow-growing mycobacteria is the lack of sequence conservation within the leader sequences, even between closely related organisms, while the coding regions are almost perfectly conserved among widely diverging species. It is immediately apparent that the length of the leader region for the *rrn* operons of slow-growing mycobacteria studied is variable from 183 bp for *M.paratuberculosis* to 207 bp for *M.leprae* (Table 4.2). The sequence alignments in **Fig.4.2** show that the regions of similar sequence are confined to certain domains within the leaders. For example, the Box B element which contains about twenty nucleotides at the start of transcription, and the region from positions 34-66 which includes Box A, Box C and the

Table 4.2 Comparison of properties of the leader regions of mycobacteria

Species	Leader (bases)	Sequence differences		Sequence similarity of 16S rRNA (%) ^c
		leader region ^a	16S rRNA (helix10) ^b	
<i>M. avium</i>	187	0	0	100
<i>M. paratuberculosis</i>	183	4	0	99.9
<i>M. intracellulare</i>	189	10	6	99.4
<i>M. lufu</i>	188	16	2	*
<i>M. simiae</i> (<i>M. habana</i>)	205	44	6	96.3
<i>M. marinum</i>	198	45	6	98.6
<i>M. leprae</i>	207	64	4	>98.5
<i>M. tuberculosis</i>	191	82	5	98.6

*. The sequence of the 16S rRNA gene is not available.

a. See **Figure 4.2**. b. See **Table 4.1**. c. See Rogall et al., 1990b.

RNase III processing sites, are identical among all the species. Downstream from the conserved section of 32 nucleotides to the 5'-end of the 16S rRNA gene, the sequences show more differences. The discrepancy is greater than sequence differences in the V2 region of the mature 16S rRNA gene (see **Table 4.2**), the most variable region of the 16S rRNA gene.

4.4) Discussion

A characteristic of *E.coli* *rrn* operons is the presence of the functional elements responsible for the stringent and growth-rate controls and the efficient transcription of the operons. Two tandem promoters are found in each *rrn* operon, of which the P1 is most active during rapid cell growth, and is the only one subject to stringent control. The Box A sequence downstream of promoter P2 is required for efficient rRNA chain elongation (Gourse et al., 1986). Since only one promoter is found in the *rrn* operon of slow-growing mycobacteria, it is predictable that whole functional elements can be found in the single *rrn* operon.

The evidence that *M.tuberculosis* maintains growth-rate-dependent control of ribosomal biosynthesis rests on the observation that the RNA:DNA ratio varies according to growth conditions whereas the ratio RNA:protein remains unchanged (Winder & Rooney, 1970). Although stringent control has not been detected in mycobacterial species because of the relatively small changes in amounts of rRNA that occur in these slow-growing cells, the universal rRNA

synthesis mechanism is likely to exist in mycobacteria. Since the sequences responsible for the growth rate control and the stringent control are thought to be the same (Ryals et al., 1982ab), and the sequence responsible for stringent control can always be found near the transcription start site, it is likely that the Box B element in the leader region of *M.tuberculosis* *rrn* operon has the same functions. However, little is known of the antitermination mechanism in mycobacterial *rrn* operons. The sequence data presented above for *M.tuberculosis* and the other slow-growing mycobacteria, showing similarities with elements of the leader region of the *rrn* operons of *B.subtilis* and *E.coli*, suggest that they have similar mechanisms for the control of rRNA synthesis. The conservation of Box B elements among the slow-growing mycobacteria generally suggests that, as for *M.tuberculosis* and *M.leprae* *rrn* operons, the transcription start sites for other *rrn* operons of slow-growing mycobacteria are also located in this region.

It has been suggested that processing of rRNA molecules into mature RNA transcripts involves the generation of stem-loop structures between complementary sequences flanking the mature transcripts (Robertson et al., 1967). These structures serve as processing signals, for example in *B.subtilis*, for specific RNases (Loughney et al., 1983). Sequence analysis of the leader regions of slow-growing mycobacteria revealed a region of 32 bp (35-66; **Fig.4.2**) with striking similarity to the processing signals located in the 5'-end of 16S rRNA genes of

Table 4.3 Comparison of invariant leader regions of mycobacterial *rrn* operons with those of other species.

Species	Nucleotide Sequence
GRAM-POSITIVE	
<i>Actinomycetes</i>	
<i>Mycobacteria</i> ^a	
<i>Streptomyces</i> ^b	
	<div style="display: flex; justify-content: space-around;"> ←Box A→ ←Box C→ </div>
	5'-TGTTGTTTGA GAA↓CT↓CAATA GTGTGTTTGG T-3' 5'-CGTTCCTTGA GAA CT CAACA GCGTGCCAAA A-3'
<i>Bacilli</i>	
<i>B. subtilis</i> ^c	5'-AGTTCTTTGA AAA CT AAACA AGACAAAACG T-3'
<i>Mycoplasma</i>	
<i>M. capricolum</i> ^d	5'-CGATCTTTGA AAA CT AAATA GAATAATTAT T-3'
<i>M. hyopneumoniae</i> ^e	5'-AGATCTTTCA AAA CT AGGTA TATAATAAAA A-3'
GRAM-NEGATIVE	
<i>E. coli</i> ^f	5'-TGCTCTTTAA CA N ₁₅ TGTGTGGGCA C-3'

a. See **Figure 4.2**; b. Wezel et al., 1991 and Pernodet et al., 1989; c. Ogasawara et al., 1983; d. Iwami et al., 1984; e. Taschke & Herrmann, 1986; f. Li et al., 1984; ↓. The arrows indicate RNase III cleavage sites in *B. subtilis* (Ogasawara et al., 1983).

B.subtilis, *Mycoplasma capricolum*, *Mycoplasma hyopneumoniae* and *Streptomyces coelicolor* (see **Table 4.3**). The data suggest that slow-growing mycobacteria may possess rRNA processing mechanisms similar to those found in *B.subtilis*, *Mycoplasma* and *Streptomyces*. Since this conserved region contains both Box A element and RNase III processing sites, it is likely that this region has a dual activity. During transcription initiation it serves as a site for antitermination, whereas when the gene has been transcribed, the stem-loop structure can form in the transcripts and serve as a processing site.

The methods used in this chapter make the leader regions of the *rrn* operons of mycobacteria as readily available for study as the 16S rRNA sequences. The greater variability of the sequences of the leader regions compared with the sequences of 16S rRNA genes, make the leader region a more suitable area for a study of phylogeny. The summary of the data obtained (see **Table 4.2**) further supports the notion that *M.avium* and *M.paratuberculosis* are very closely related, not only to each other but also to *M.intracellulare* (Rogall et al, 1990ab). The data of Fig. 4.2 and Table 4.2 also show that *M.lufu* is another close relative of *M.avium*. Pre-16S rRNAs of *M.marinum* and *M.leprae* conform to the general pattern. Although *M.leprae* is an extremely slow-grower, it has a typical mycobacterial *rrn* operon (Sela et al, . 1989). The interpretation of the data for *M.habana* is less straight forward. On the basis of biochemical and serological tests, Meissner &

Schroder(1975) concluded that *M.habana* strains belong to the species *Mycobacterium simiae*. This conclusion is supported by our sequence data, for part of the 16S rRNA coding region of *M.habana* is identical with that obtained for *M.simiae* by Rogall et al (1990a) (see for example **Table 4.1**). The sequence of the leader region of *M.simiae* suggests that *M.simiae* is more closely related to *M.avium* than is *M.leprae*. In contrast, the data for *M.simiae* 16S rRNA sequences place *M.simiae* as the most distant slow-growing relative of *M.avium* (Rogall et al., 1990b). The data for *M.tuberculosis* reveals an interesting paradox. According to the 16S rRNA data *M.marinum* and *M.tuberculosis* are as closely related (99.4% sequence identity), as are *M.avium* and *M.intracellulare* (see **Table 4.2**). However, the supplementary data reveal a different situation. In total there are 10 nucleotide differences in the leader regions of *M.avium* and *M.intracellulare* but many more (82) differences between *M.marinum* and *M.tuberculosis*. These differences are spread throughout the region (**Fig.4.2**).

CHAPTER FIVE

Internal Transcribed Spacer of 16S-23S rRNA Genes of Slow-Growing Mycobacteria

5.1) Introduction

In *E.coli* *rrn* operons the internal transcribed spacer (ITS) between 16S and 23S rRNA genes is found to form a complementary stem-loop structure with the sequence from the leader region. It is involved in rRNA processing after the *rrn* operons have been transcribed (Apirion & Miczak, 1993). This region also contains antitermination elements that are similar to the *rrn* leader region Box A sequences, preceded by Box B-like dyad symmetry (Berg et al., 1989). Little is known about the ITS area of mycobacteria, except that unlike in *E.coli* there is no tRNA gene found in the region. Knowledge of the mycobacterial ITS sequences would shed some light on the function of this area. A secondary consideration is that the existing variable regions in mycobacterial ITS can be explored to differentiate closely related species. Many investigations have been focused on the 16S rRNA gene and have tried to use the sequence to differentiate bacteria at the genus and species level. However, 16S rRNA gene sequences do not vary greatly within a species as, for example, in the *M.tuberculosis* complex. The study on the leader region of slow-growing mycobacteria in Chapter 4 has shown that it is a potential area for

differentiation of closely related species. Another likely area is the ITS. The ITS region should be under less selective pressure during evolution than the rRNA genes themselves and therefore should vary more extensively than sequences within genes that have functional roles.

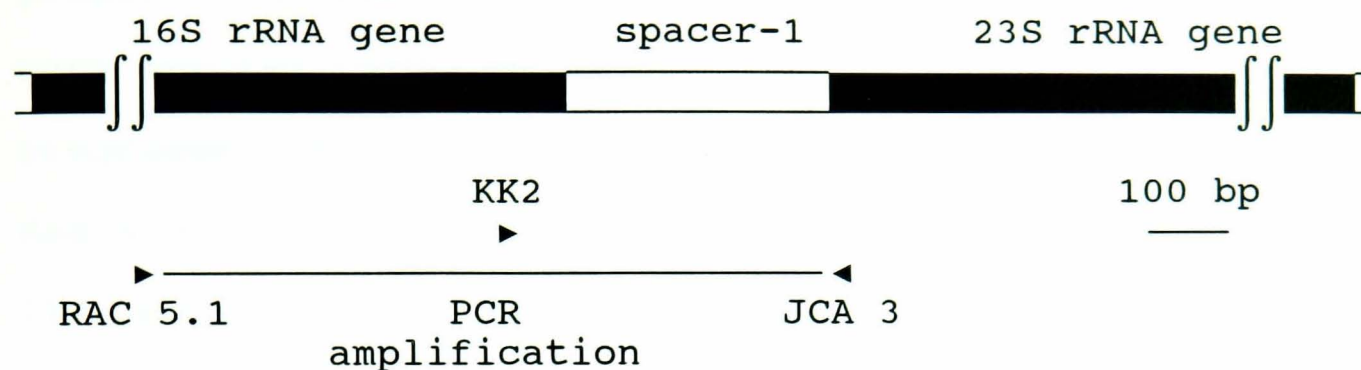
5.2) Results

5.2.1) The internal transcribed spacer of the 16S-23S rRNA spacer sequence

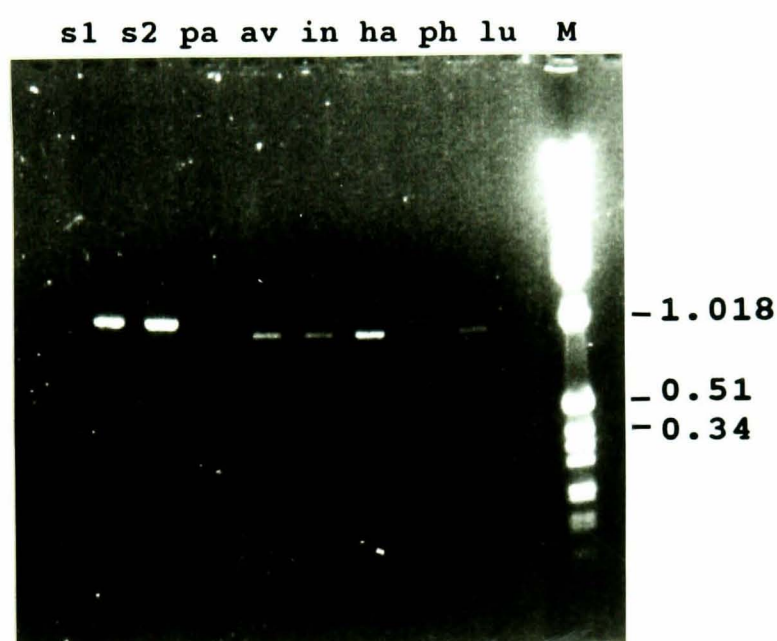
The cloning strategy is shown in Fig.5.1. The existence of a single *rrn* operon in slow-growing mycobacteria makes the ITS area available for PCR amplification, yielding a single product. To clone the ITS, PCR amplification was applied using two primers made from sequences which are conserved among prokaryotic 16S and 23S gene coding regions. The following primers were used to amplify the 16S-23S rDNA ITS: RAC5.1 which was made from the 16S rRNA gene coding region with the sequence 5'-AATTGACGGGGGCCCGCACAA-3'; and JCA3 which was made from the 5'-end of 23S rRNA of *M.leprae* with the sequence 5'-GCCAAGGCATCCACCATGC-3'. Their positions within the *rrn* operon are shown in Fig.5.1. The PCR experiments were carried out as described in Materials and Methods. PCR products were visualized on 1% agarose gels stained with ethidium bromide (Fig.5.1). After purifying DNA from the agarose gel by the glass milk binding method (see Materials and Methods), the PCR products were then cloned into

Fig.5.1 Amplification of the ITS regions of mycobacteria

A



B



C

Primer RAC5.1 5' AAT TGA CGG GGG CCC GCA CAA 3'

JCA3 5' GCC AAG GCA TCC ACC ATG C 3'

KK2 5' AAG TCG TAA CAA GGT AGC CGT ACC 3'

(A) Polymerase chain reaction strategy to amplify the ITS region between 16S and 23S rRNA genes of slow-growing mycobacteria. (B) PCR products were run on 1% agarose gels and stained with ethidium bromide. (C) Primers used in PCR amplification and sequencing. **s1**- *M.smegmatis* NCTC8159; **s2**- *M.smegmatis* mc²155; **pa**- *M.paratuberculosis*; **av**- *M.avium*; **in**- *intracellulare*; **ha**- *M.habana*; **ph**- *M.phlei*; **lu**- *M.lufu*; **M**- Lambda 1 kbp ladder markers, size given in kbp.

pBluescript vector restricted with *EcoRV*. The clone containing each PCR product insert was sequenced with primers KK2 and JCA3 on both strands. The sequences for each species were established and aligned using the PILEUP programme of GCG V.7 (Genetics Computer Group Inc., Madison, Wisconsin) (see **Fig.5.2**). Previously established ITS sequences for *M.leprae* (Liesack et al., 1991) and *M.tuberculosis* (Kempell et al., 1992) were also included in the alignment.

5.2.2) Identification of functional elements in the ITS

The six sequenced mycobacterial *rrn* operon 16S-23S spacer regions contain Box A sequences that are similar to the *rrn* leader region Box A sequences, except for a single base difference at position three. In each of these spacer regions, Box A is preceded by a Box B-like dyad symmetry with nearly identical primary sequences. No Box C element is found in this region. In *E.coli*, according to Berg et al. (1989) the *rrn* operon antitermination mechanism is also invoked by the spacer region. From sequence homology, this mechanism appears to exist in the mycobacterial *rrn* operon ITS as well.

The first seventeen nucleotides at the 5'-end of ITS show completely identical sequences containing two purine-rich tracts. The first tract at the 5'-terminus has the potential to interact with the anti-Shine-Dalgarno sequence (Shine & Dalgarno, 1974) at the 3'-end of 16S rRNA; and the second tract has the potential to interact with the 5'-end

S50

M.in. AAGGAGCACCACGAAAAGCACTCCAATT•G•GTGGGGTGCGAGCCGTGAG
M.av. AAGGAGCACCACGAAAAGCACcCCAaCT•G•GTGGGGTGCGAGCCGTGAG
M.lu. AAGGAGCACCACGAAAAGCtaccCAATT•G•GTGGGGTGCGAGCCGTGAG
M.si. AAGGAGCACCACGAgAAaCACTCCAATTgG•GTGGGGTGtGAGCCGTGAG
M.le. AAGGAGCACCACGAAAAaCACTCtAAaTaGttTaGGGTGtaAGCCGTcAG
M.tb. AAGGAGCACCAGcAAAAcgcCcCaAcT•••gGTGGGGcGtagGCCGTGAG

S100

M.in. GGGTTCCCGTCTGTAGTGGACGGGGGCCGGGTGCACAACAGCAAATGATT
M.av. GGGTTCCCGTCTGTAGTGGACGGGGGCCGGGTGCgCAACAGCAAATGATT
M.lu. GGGTTcTcGTCTGTAGTGGACGaaGaCCGGGTGCACAACAGCAAATGATT
M.si. GGGTTcTcGTCTGTAGTGGACGaGGGCCGGGTGCACAACAaCAggcaATc
M.le. GGGTTcTcCaTCTGTAGTGGAtGaGaGCCGGGTGCgCAACAGCAA•••ATa
M.tb. GGGTTcTtGTCTGTAGTGGgCGaGaGCCGGGTGCAtgACAaCAAagtgg

← stem C (invariant) →

S150

<i>M.in.</i>	GC•CAGACACACTATTGGGCCCTGAGACAACAC	TCGG•••••TCGATCC
<i>M.av.</i>	GC•CAGACACACTATTGGGCCCTGAGACAACAC	TCGG•••••TCcgTCC
<i>M.lu.</i>	GC•CAGACACACTATTGGGCCCTGAGACAACAC	TCGG•••••TCGAgCC
<i>M.si.</i>	GC•CAGACACACTATTGGGCCCTGAGACAACAC	TCGGccgactTCGgTtg
<i>M.le.</i>	tC•CAGACACACTgTTGGGtCCTGAGgCAACAC	TCGccttgatT•••••
<i>M.tb.</i>	cCaCcaACACACTgTTGGGtCCTGAGgCAACAC	TCGG•••••acTtGtTCC

		Box B		Box A	S200
<i>M.in.</i>	GTGTGGAGTCCCT	CCATCTTGGTGGT•GG	GGTG	TGGTGTTT	GAGTATTGG
<i>M.av.</i>	GTGTGGAGTCCCT	CCATCTTGGTGGT•GG	GGTG	TGGTGTTT	GAGTATTGG
<i>M.lu.</i>	GTGTGGAGTCCCT	CCATCTTGGTGGT•GG	GGTG	TGGTGTTT	GAGTATTGG
<i>M.si.</i>	aaGTGGtGTCCCT	CCATCTTGGTGGT•GG	GGTG	TGGTGTTT	GAGTATTGG
<i>M.le.</i>	GgGTGttGTCCCT	CatctTTGGTGGT•G	GGGTG	TGGTGTTT	GAGTATTGG
<i>M.tb.</i>	agGTGttGTC	CCaCCgcCTTGGTGGTgG	GGGTG	TGGTGTTT	GAGaAcTGG

S250

	Box C	
<i>M.in.</i>	ATA GTGGTTG	CGAGCATCTAGATGAGCGCATAGTC•••••CTTAGGGCTG
<i>M.av.</i>	ATA GTGGTTG	CGAGCATCTAGATGAGCGCATgGTC•••••tTcgtGGCcG
<i>M.lu.</i>	ATA GTGGTTG	CGAGctaCTAGATGAaCGCgTAGTC•••••CTTgGGGCTG
<i>M.si.</i>	ATA GTGGTTG	CGAGCATCTAaATGAaCGCgTcGcC•••••ggcAacGgTt
<i>M.le.</i>	ATA GTGGTTG	CGAGCATCTAaATGgatGCgTtGTCagttatgTAGtGgTG
<i>M.tb.</i>	ATA GTGGTTG	CGAGCATC•••••AatGgATAcgCtgccggcTAGcGgTG

S296

M.in. ATG•CGTCGTCG•AAATGTGTAATTTCTTCTTTGGTTTTTGTGTGT
M.av. gcG••tTCaTCG•AAATGTGTAATTTCTTtTTTaaCTcTTGTGTGT
M.lu. AcGagtTCaTCG•AAATGTGTTtATTTCTTtTTTaaCTcTTGTGTGT
M.si. AcGtgtTCGTtt•tgtgtaaTttTTTCTTtTggttTTTgTGT•••••
M.le. gcGtatTCaTtGaAAATGTGTAATTTtTtTtGgTTTTGTGTGT
M.tb. gcGtgtTctTtGtgCAatatTctTTggTTtTTgttgTgT•••••

of 16S rRNA. Thus it is possible that these first twenty nucleotides of ITS play a key role in the processing of pre-rRNA.

Like the *rrn* operons in *E.coli*, the mycobacterial ITS DNA sequences, together with the sequences for the leader region, predict that the regions that flank the 5'- and 3'-ends of mature 16S in the 30S rRNA precursor can be extensively base paired. The putative stem formed by interaction between the leader and the ITS regions is stabilized by fifty or more base-pairs (see **Fig.6.1** and **6.2**). The ITS sequence complementary to the leader sequences includes a highly conserved region of thirty-three nucleotides, with twenty-six positions being the same in all the species studied (see **Fig.5.2**). This conserved region is believed to be involved in processing of the 30S pre-RNA by RNase III, as with the *E.coli* ITS regions (Young & Steitz, 1978).

5.2.3) Greater sequence variability than 16S rRNA genes

The established ITS sequences for the four species of slow-growing mycobacteria, together with previously established ITS sequences for *M.leprae* and *M.tuberculosis*, showed length differences (see **Table 5.1**). However, compared with about 440 nucleotides for the *E.coli* ITS, which contains a tRNA gene (Young et al., 1979; Harvey et al., 1988), the presence of only approximately 280 nucleotides in the mycobacterial 16S-23S ITS region and no sequence homology to bacterial tRNA genes further confirms

Table 5.1 Comparison of properties of the ITS regions of mycobacteria

Species	length of the ITS region (bases)	Sequence differences		
		ITS region	leader region	16S rRNA (helix 10)
<i>M.avium</i>	278	0	0	0
<i>M.intracellulare</i>	279	21	10	6
<i>M.lufu</i>	280	24	16	2
<i>M.simiae</i> (<i>M.habana</i>)	283	62	44	6
<i>M.leprae</i>	285	64	64	4
<i>M.tuberculosis</i>	276	88	82	5

that there is no tRNA gene in this area.

The ITS sequence alignment in Fig.5.2 reveals some variable regions. Taking the *M.avium* ITS sequence as standard, the sequence difference with *M.intracellulare* is twenty-one nucleotides, with *M.lufu* twenty-four nucleotides, with *M.simiae* sixty-two nucleotides, with *M.leprae* sixty-four nucleotides and with *M.tuberculosis* as much as eighty-eight nucleotides (see **Table 5.1**). The most variable area is located at the 3'-end of the ITS region. Since there are only about 280 nucleotides in the ITS region, the percentages of sequence variability for the species studied range from 7.5% for *M.intracellulare* to 31.4% for *M.tuberculosis*. This variability is much greater than the 2% differences in the 16S rRNA genes of mycobacteria (Rogall et al., 1990b). This fact makes the ITS region a suitable target for species specific probes.

5.3) Discussion

The putative Box A and Box B elements identified in the ITS region of the mycobacterial *rrn* operon, suggest that the *rrn* antitermination mechanism is also invoked by the spacer region. This mechanism guarantees that the rRNA transcripts are transcribed to the end of the operon without interruption. Failure to do so might be fatal for the slow-growing mycobacteria because the possession of a single *rrn* operon leaves them without the material basis for compensation.

The universal structure of the double-stranded stems bracketing the mature 16S rRNA in *E.coli* and many other species, also appears to occur in slow-growing mycobacteria. This conclusion is based on the fact that a large proportion of nucleotides from the ITS region [see stem c (invariant) in Fig.5.2], have the potential to complement the stem (invariant) region of the leader sequences preceding the 16S rRNA. The conservation of these features during prokaryotic evolution suggests that they might be important for the formation or function of rRNA.

Variable regions of the 16S rRNA have been frequently used as the target for DNA probes to identify microorganisms (Cox et al., 1991). In some situations, however, there is very little sequence variation observed between the 16S rRNA genes of closely related microorganisms. The single *rrn* operon of slow-growing mycobacteria with the 16S-23S rRNA ITS located centrally, make the ITS accessible and, by virtue of its small size, easy to sequence. These advantages make the ITS region generally useful for mapping phylogenetic relationships among closely related species. Together with the sequences in the leader region (see Chapter 4), the ITS sequences supplement the sequence data for 16S rRNA. If the 16S rRNA sequences are taken to represent the hour hand of a clock measuring evolutionary time, then the leader and spacer regions can be thought of as the minute hand. The established ITS sequences confirm that this strategy is helpful in differentiating some closely related slow-

growing mycobacterial species. It also confirms the phylogenetic relationships noted in the study of leader regions in Chapter 4, namely that *M.avium* and *M.intracellulare* are closely related to each other, and that *M.simiae* is more closely related to *M.avium* than was suggested by Rogall et al. (1990b), based on the 16S rRNA gene sequence. *M.tuberculosis*, with as much as 88 nucleotides differences in ITS compared with the same area of *M.avium*, places this species far away in evolutionary terms from other slow-growing mycobacteria.

CHAPTER SIX

Putative Secondary Structure of Pre-16S rRNA of Slow-Growing Mycobacteria and the inferred RNase Processing Pathway

6.1) Introduction

A possible relationship between RNA secondary structure and its biological function has been demonstrated in several cases (Zaug et al., 1986; Eperon et al., 1986). Molecules with the same function have the potential to fold into very similar secondary structures, indicating the importance of those secondary structures for function. The rRNA molecules are an example (Noller, 1984). Since these structures are mainly determined by the primary nucleotide sequences, the results described in Chapter 4 and Chapter 5 provide the opportunity to work out the putative secondary structure for pre-16S rRNA of slow-growing mycobacteria.

As yet it has proved difficult to apply computer predictions of secondary structure, based on thermodynamic considerations of helix stability, to large molecules such as rRNAs. The models of secondary structure for the rRNAs have generally been derived from a combination of comparative sequence studies together with direct experimental analysis (Brimacombe et al., 1983; Noller, 1984). The sequence data from the leader region of the *rrn* operon of eight slow-growing mycobacterial species,

together with the sequence data of the ITS region from six slow-growing mycobacterial species, enabled me to establish a scheme for the secondary structures of pre-16S rRNA. The scheme predicts where secondary structure features can be accommodated into pre-rRNA of many different species. Especially useful for the analysis are situations where there is poor conservation of primary structure, but the predicted secondary structure is conserved as the result of compensatory base changes. By providing information on likely secondary structure, this scheme can also provide insights into the possible function of pre-rRNA molecules.

6.2) Results and discussion

According to the standards above, a scheme was used to establish the secondary structures of pre-16S rRNAs of slow-growing mycobacteria. The derived structures for *M.leprae* and *M.tuberculosis*, respectively, are shown in Fig.6.1 and Fig.6.2. Since the leader region sequence of *M.leprae* is similar to that of the other species studied (Fig.4.2), its secondary structure is used to represent all the species studied.

The figures show the following characteristics:

- (1) About 65% of the leader sequence forms a stable double helical structure with the ITS region, or forms helices within itself.
- (2) The remarkable conservation in the nucleotide sequences (see Fig.4.2 and Fig.5.2) of the hydrogen bonded stem formed

by interaction between the leader and ITS sequences of *rrn* operons of slow-growing mycobacteria suggests that this region is involved in processing of the mature 16S rRNA.

(3) A Box A anti-termination signal, which is identified by comparison with those found in *E.coli* and *B.subtilis* (see **Table 4.3**), also exists in this region. Box B, a sequence of dyad symmetry located 5' to Box A, is present in both species.

(4) The leader sequence of *M.tuberculosis* downstream from nucleotide 147 (see **Fig.6.2**) diverges from the sequences of the other slow-growers studied to an extent that is reflected in its secondary structure. Instead of forming five helices in the leader region, *M.tuberculosis* forms only three.

(5) Sequences in the hairpin loop region of helix L1 and helix S4 are complementary and therefore have the potential to interact.

(6) the 18 nucleotides at the 5'-end of the ITS region have the ability to bring the 5'- and 3'-ends of the 16S rRNA close together to form helix S1, as shown in **Fig.6.1** and **Fig.6.2**.

Pre-16S rRNA is likely to be produced by RNase III which introduces a double cleavage in the stem (see **Figs.6.1** and **6.2**). Taking *M.leprae* as an example, pre-16S rRNA is likely to comprise the 16S rRNA together with approximately 160 nucleotides upstream from the 5'-end and approximately 113 nucleotides downstream from the 3'-end. Two other RNases are implicated in the maturation process;

Figure 6.1 The secondary structure of *M.leprae* pre-16S rRNA.

The putative secondary structure of part of the transcript of the *rrn* operon of *M.leprae*. The scheme is based on the data of **Fig.4.2** and **Fig.5.2**. The 5'- and 3'-ends of the 16S rRNA are shown by arrows and the 16S rRNA sequences are boxed. Part of the loop of helix L1 and of helix S4 are boxed to indicate a possible interaction between the two motifs in the tertiary structure. The capacity of the 5'-terminal sequence of spacer-1 to interact with both the 5'- and 3'-ends of the 16S rRNA is illustrated. L10 etc., refer to the leader region starting from the consensus 5'-end of pre-rRNA. S10 etc., refer to the spacer sequences starting from the 3'-end of mature 16S rRNA (see also **Fig.5.2**). The arrows between residues L40 and L50, and between S115 and S120, indicate RNase III cleavage sites. The arrow following S130 indicates the continuation of the polynucleotide chain.

M. leprae

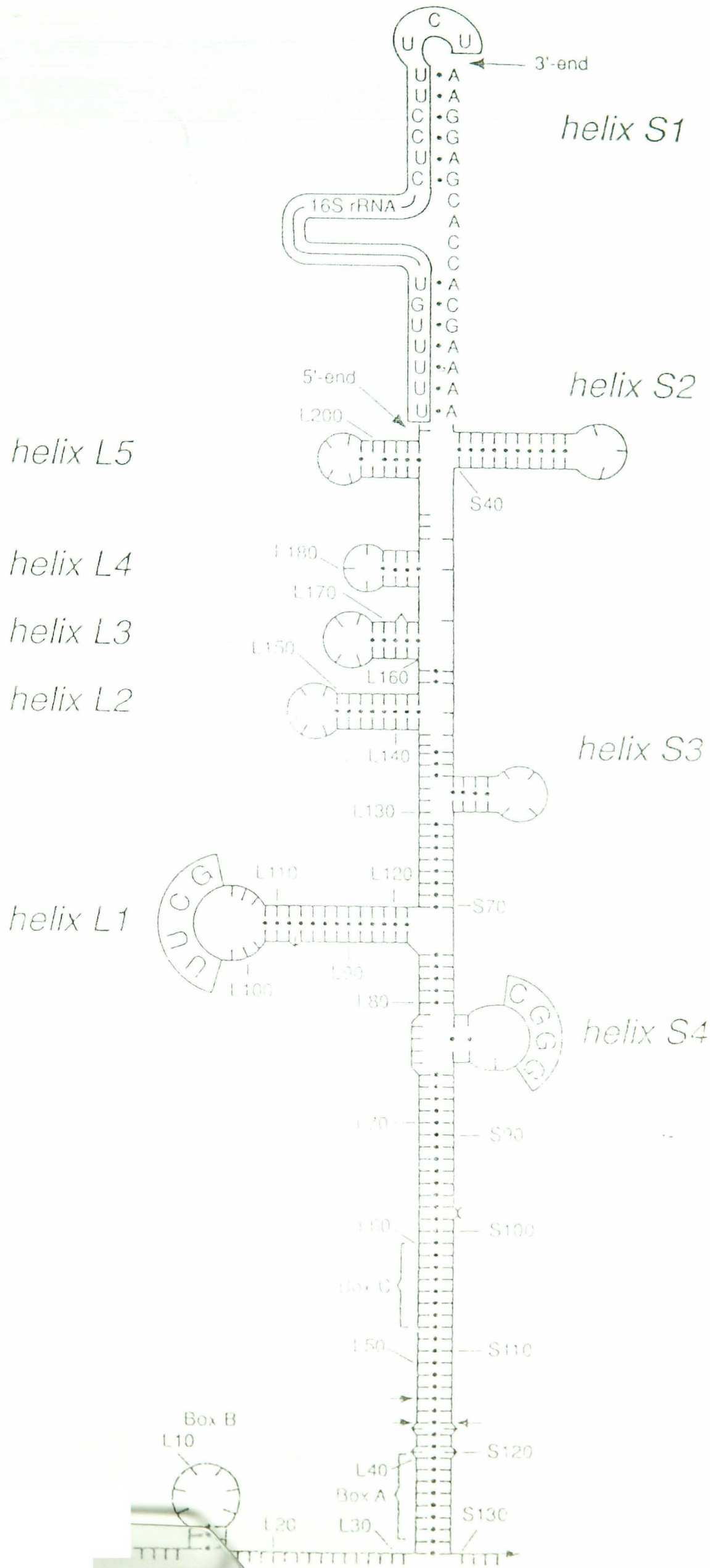


Figure 6.2 The secondary structure of *M.tuberculosis* pre-16S rRNA.

A possible secondary structure of part of the transcript of the *rrn* operon of *M.tuberculosis*. The legend is the same as that in **Fig.6.1**, except that there are three helices in the leader region of *M.tuberculosis rrn* operon, two fewer than that in other slow-growing mycobacterial *rrn* operon.

M.tuberculosis

helix L3

helix L2

helix L1

L110

Box B

L10

The diagram illustrates the structural organization of the M. tuberculosis protein. It features several helices labeled L1, L2, L3, and L10, and a region labeled Box B. Helix L1 is shown as a semi-circular arc with four internal loops, labeled L110. Helix L2 is a straight line segment. Helix L3 is a straight line segment. Helix L10 is a semi-circular arc with four internal loops, labeled L10. Box B is a rectangular region located between helices L1 and L10. The helices are arranged in a specific spatial relationship, with L1 and L10 forming a semi-circular shape around the central region.



viz, one which cleaves R_pU to generate the 5'-end, and another which cleaves U_pA to generate the 3'-end of the 16S rRNA. This pathway, leading to mature 16S rRNA, is in accord with that proposed for *E.coli* (for review see Apirion & Miczak, 1993). However, except for RNaseIII, little is known of the enzymes involved.

The mechanism of generation of the 5'-end of 16S rRNA is inferred as follows. The potential 5'-end of 16S rRNA is marked by helix L5, helix S1 and helix S2, as illustrated in Fig.6.1. In each case there is either a G_pU or an A_pU bond in which the G or A residue is not base-paired but the U residue is paired with an A residue in the first base-pair in a run of either six or seven base-pairs. It is proposed that the 5'-end of 16S rRNA is generated by the cleavage of this R_pU bond. The predicted 5'-end of 16S rRNA is $_pUUUUUGU...$ for *M.leprae*, *M.marinum*, *M.lufu* and *M.intracellulare*, and is $_pUUUUGU$ for *M.tuberculosis*, *M.bovis*, *M.avium*, *M.paratuberculosis* and *M.simiae*. The 5'-end of *M.bovis* 16S rRNA, determined by the primer extension method (Suzuki et al., 1988), was reported as $_pUUUGU$. The difference of one U residue in the observed and predicted sequence is within experimental error.

The mechanism of generation of the 3'-end of 16S rRNA is inferred as follows. The sequence of the anti-Shine-Dalgarno element at the 3'-end of 16S rRNA and the contiguous purine-rich tract at the 5'-end of the ITS region, are the same in all the slow-growers studied. The two regions have the capability of interacting to form a

helical loop as shown in Figs.6.1 and 6.2 respectively for *M.leprae* and *M.tuberculosis*. According to this scheme, the 3'-end of 16S rRNA of the slow-growers is generated by the cleavage of a U_pA bond, the U residue being located within a loop of non-base-paired residues.

CHAPTER SEVEN

Pre-rRNA of Fast-Growing Mycobacteria as Exemplified by

M.smegmatis NCTC 8159

7.1) Introduction

The sequences of the flanking region of the 16S rRNA gene have been shown to be highly conserved in slow-growing mycobacteria, demonstrating considerable homology (see Chapter 3 and Chapter 4). The occurrence of only a single promoter in the *rrn* operon of slow-growing mycobacteria makes this operon different from those described in other eubacteria.

The existence of the two *rrn* operons in fast-growing mycobacteria (Bercovier et al., 1986), and the published physical maps of the two sets of rRNA genes of *M.smegmatis* (Bercovier et al., 1989; also see Fig.7.3 A), which are defined as *rrnI* and *rrnII* operons in this study, provided an opportunity to look into the flanking sequences of the 16S rRNA genes of *M.smegmatis*. In this chapter, I present the following data relating to the two *rrn* operons of *M.smegmatis*: (1) cloning and sequencing of the region upstream from the 5'-end of the 16S rRNA gene; (2) cloning and sequencing of the ITS region; (3) the identification of rRNA transcripts and their processing sites.

7.2) Results

7.2.1) Cloning and sequencing of the upstream region of the first 16S rRNA gene of *M.smegmatis* NCTC 8159

Genomic DNA of *M.smegmatis* NCTC 8159 was isolated (see Materials and Methods) and used as the substrate for PCR amplification. Primers PL1 and P4 were used in PCR and their positions are shown in Fig.4.1. After 35 reaction cycles (see Materials and Methods), the reaction mixture was loaded on to a 1% agarose gel and a single band of about 580 bp was detected after electrophoresis (Fig.4.1). The band was cut from the gel and the DNA was purified by the glass milk binding method (see Materials and Methods). The purified DNA was then cloned into Bluescript derived T-Vector (see Materials and Methods). The recombinant DNA was then used to transform *E.coli* DH5 α F' strain, and the clones identified by blue and white selection. Primers used in establishing the sequence are shown in Fig.4.1. The resultant sequence is given in Fig.7.1: top line.

The single band product obtained using PL1 and P4 primers might have been produced from either one or both *M.smegmatis* *rrn* operons. To investigate whether the conserved Box B-like element only exists in one *rrn* operon of *M.smegmatis*, but not in the other, the following experiment was performed, based on the restriction map of the two *rrn* operons of *M.smegmatis* (Bercovier et al., 1989, also see Fig.7.3). The *Pst*I enzyme digested genomic DNA of *M.smegmatis* was run on a 1% agarose gel, then sections of the gel around 1.4 kbp and 3.2 kbp were removed and the

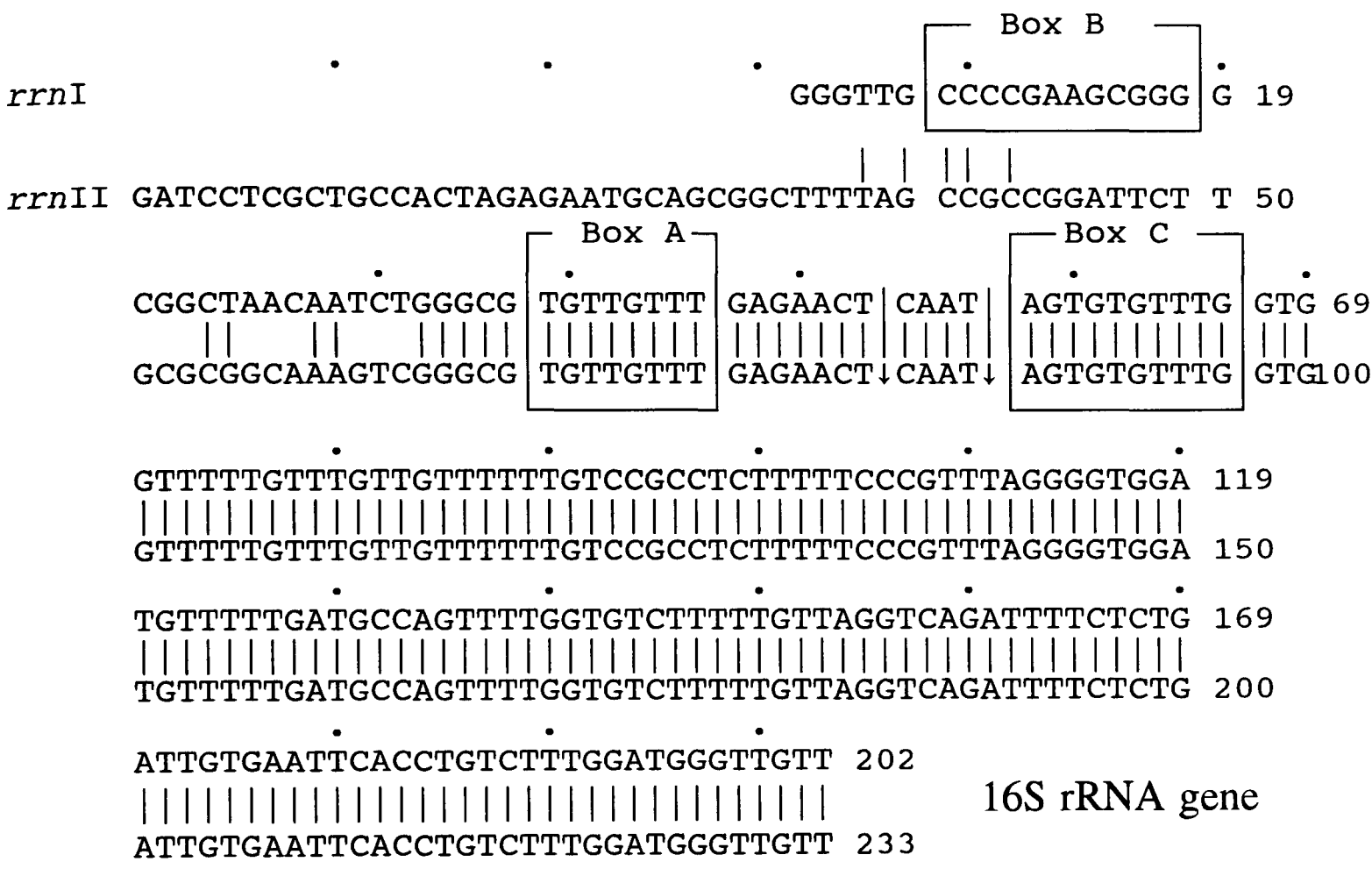


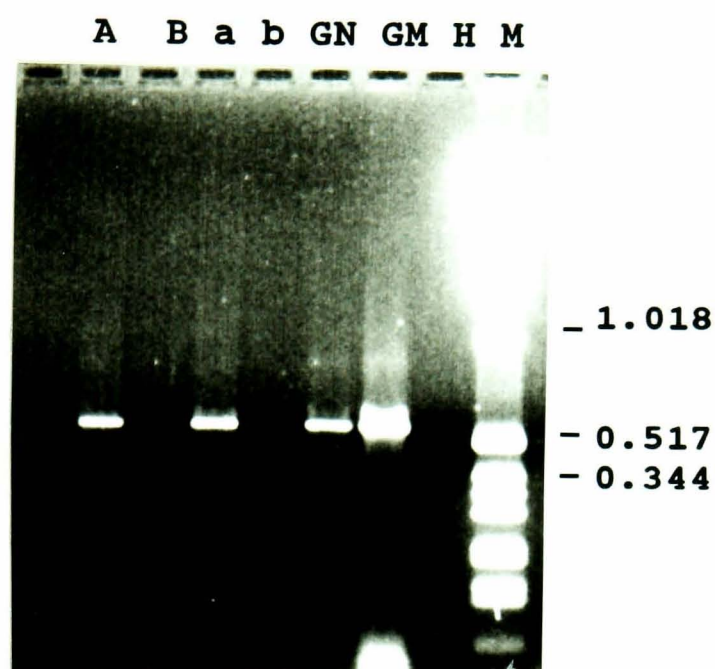
Figure 7.1 The sequences upstream from the 16S rRNA genes of *M. smegmatis*. The *rrnI* and *rrnII* operons are shown. The putative Box A, Box B and Box C elements are boxed. There is no conserved Box B element in the *rrnII* operon. The arrows indicate RNase III processing sites.

restricted DNAs purified by the glass milk binding method (see Materials and Methods). The purified DNAs were used as substrates, with PL1 (made from the Box B element) and P4 as primers for PCR amplification. The result is shown in Fig.7.2. Only restricted DNAs around the size 1.4 Kbp produced a band, which implied that only *rrnI* (see Fig.7.1) contained the Box B element, and *rrnII* did not. Since all the slow-growing mycobacteria studied contain the Box B element in their *rrn* operons, it is apparent that there is a slow-grower-like *rrn* operon in *M.smegmatis*.

7.2.2) Cloning and sequencing of the upstream region of the second *rrn* operon of *M.smegmatis* NCTC 8159

To understand the control mechanism for rRNA genes of *M.smegmatis*, an attempt was made to clone the leader and promoter regions of the second *rrn* operon. The strategy of inverse PCR was used to obtain the two 5'-flanking regions. This strategy is illustrated in Fig.7.3. The genomic DNA of *M.smegmatis* NCTC 8159 was digested with *Bam*H1 and then self ligated at low DNA concentration in order to circularize the restricted fragment (see Materials and Methods). The circularized DNA fragments were used as substrates for PCR. Based on the restriction map of Fig.7.3 (A), two PCR products with the size of 1.3 kbp and 1.0 kbp were expected. A 1.0 kbp PCR product of the *rrnII* operon was in fact obtained (see Fig.7.3 C) and the band was cut from the agarose gel, purified by means of glass milk binding, and cloned into T-vector (see Materials and Methods). The

Fig.7.2 *rrnI* operon of *M.smegmatis* contains a slow-grower-like Box B element



Primers used: PL1 see **Figure 4.1**
P₄ see **Figure 4.1**

Polymerase chain reaction experiment to confirm that the Box B element only exists in the *rrnI* operon of *M.smegmatis*, and not in the *rrnII* operon. **A** and **a**- DNA substrates of NCTC8159 and mc²155 strains respectively, isolated from *Pst*I digested 1.4 kbp gel sections; **B** and **b**- DNA substrates of NCTC8159 and mc²155 strains respectively, isolated from *Pst*I digested 3.2 kbp gel section; **G.N**- genomic DNA of NCTC 8159 strain; **G.M**- genomic DNA of mc²155 strain; and **H**- water as control; **M**- Lambda 1 kbp ladder markers, size given in kbp.

A

16S	23S	5S
-----	-----	----

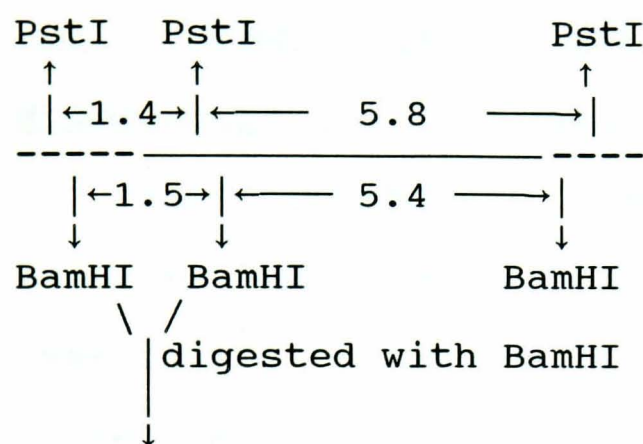
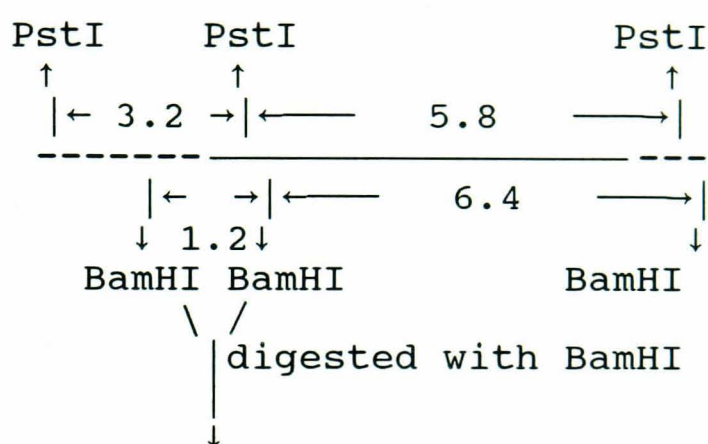
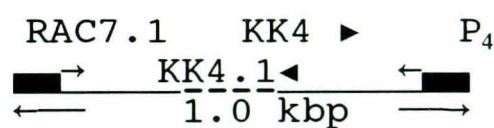
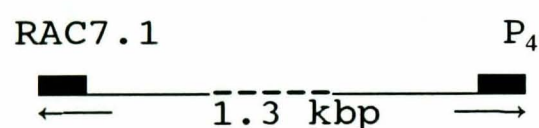
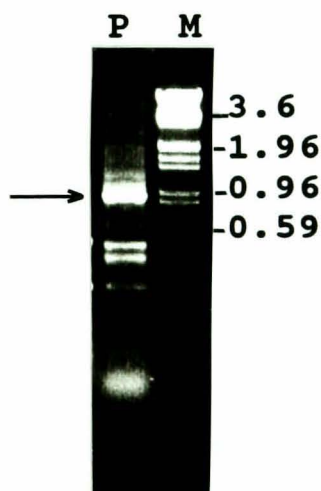
I**II****B****C**

Figure 7.3 The strategy of inverse PCR.

(A) The inverse PCR strategy used to clone the leader region of the second *rrn* operon of *M. smegmatis*. The arrows shown in (B) were primers used in the sequencing; (C) PCR gel showing the 1.0 kbp band obtained by inverse PCR. **P**-inverse PCR products, the 1 kbp band is indicated by arrow; **M**- Lambda/*EcoRI*+*HindIII* digested markers, size given in kbp. The solid line indicates 16S rRNA gene and the broken line indicates the flanking region of 16S rRNA gene.

resultant plasmid recombinant was named pJY105. The 1.0 Kbp fragment was sequenced on both strands (see Materials and Methods) and the 233 bp sequence upstream of the 16S rRNA gene of the *rrnII* is shown in Fig.7.1: bottom line.

7.2.3) Comparison of the upstream sequences of the two *M.smegmatis* 16S rRNA genes

The recombinant clones obtained as described in sections 7.2.1 and 7.2.2 were sequenced. Part of the 16S rRNA coding region was established for each of the operons, as well as the upstream areas preceding the 16S rRNA genes. The helix10 of the V2 region of the mature 16S rRNA gene (Fig.3 from Kempell et al., 1992) has the same sequence in both operons, which corresponds to the published sequence for *M.smegmatis* (Rogall et al., 1990a). The comparison confirms the identity of the species studied and also indicates that the two operons probably have identical 16S rRNA coding regions.

The two sequences preceding the 16S rRNA genes are illustrated in Fig.7.1, which shows that the *rrnII* operon obtained by the inverse PCR was different from that obtained by normal PCR and upstream of the Box A element. There is no conserved Box B-like element in the *rrnII* operon. The comparison suggests that the two recombinant clones obtained were from the two different *rrn* operons of *M.smegmatis* NCTC 8159.

7.2.4) Cloning and sequencing of the ITS region between 16S and 23S rRNA genes of *M.smegmatis*

The ITS region of *M.smegmatis* was cloned and sequenced

using the strategy illustrated in Fig.5.1. Since the PCR fragments amplified from *M.smegmatis* and *M.phlei* genomic DNAs are bigger than the equivalent fragments amplified from slow-growing mycobacteria, the size of the ITS region in fast-growing mycobacteria is longer than those of slow-growers. The established sequence confirmed this point (see **Fig.7.4**). The length of the ITS for *M.smegmatis* is about 365 bp, 85 bp longer than the length of the ITS from slow-growing mycobacteria. Like the ITS region of slow-growing mycobacteria, the putative Box A and Box B elements were also present (see **Fig.7.4**), indicating that the anti-termination mechanism also exists in the ITS region of fast-growing mycobacteria.

7.2.5) The putative secondary structure of *M.smegmatis* pre-16S rRNA and its processing

The availability of sequences upstream of the 16S rRNA gene and the ITS region of *M.smegmatis*, enabled me to establish the putative secondary structure for *M.smegmatis* pre-16S rRNA molecules. The structure conforms to the scheme of secondary structure illustrated in Fig.6.1 and Fig.6.2 for pre-16S rRNA of slow-growing mycobacteria. The characteristics are as follow: (1) the first 20 nucleotides of the 5'-end of the ITS are similar to those of slow-growing mycobacteria (see **Fig.5.3**), and they have the potential to interact with the 5'- and 3'-ends of the 16S rRNA gene; (2) nucleotides from the leader and the ITS regions form a 34 nucleotide long stem structure, which stabilizes the whole secondary structure and provides a

Fig. 7.4 The sequence of *M.smegmatis* ITS region

```
1      AAGGAGCACC ACGAGAGACA CTCTCCGTTG GGGACGGTGT GAGCCGTGAG
51     GAGCTGGAGC GCTGTAGTGG CGCCGGCTTG GTGCACAGCA AACGTTGAGA
101    TGCGGTGTGG GAAACGCTGT TTCGATGGAC TGCCAGACAC ACTATTGGGC
151    CCTGAGACAA CAGGCCCGTT GTTCCCTGGC CACTGTGTGT GGTGGGAGGC

                Box B                Box A
201    GTGTTGTTGC CCTGCTTTGG TGGTGGGGTG TGGTGTTTGA TTTGTGGATA

        Box C
251    GTGGTTGCGA GCATCTAGTT CGTAAGAGTG TGGCTGCCGG CCTTTGAGGT
301    TGGGTGGCGC ATTGTTGCGG ACAAATTGAT GTGCCATTTT TCTTCTGATT
351    ATTGGTTTTT TGTGT
```

Figure 7.4 The sequence of the ITS region of *M.smegmatis*.
The putative Box A, Box B and Box C anti-termination
elements are underlined.

location for RNase III processing; (3) Helix L1 and Helix S4 have the potential to interact through complementary sequences located within the hairpin loops; (4) although the ITS region is 85 bp longer, approximately the size of a tRNA gene, no tRNA sequence was found in this region. However, the Helix S4, in the scheme of secondary structure of *M.smegmatis* pre-rRNA, is much longer (see **Fig.7.5**) than Helix S4 of slow growers.

The assumption that, like the pre-rRNA of fast-growing bacteria in general, mycobacterial pre-rRNA is processed by RNaseIII at the stem formed by the interaction between the leader and the ITS regions, was investigated by primer extension experiments. The relative abundance of rRNA in *M.smegmatis* made the experiment possible. Total RNA from *M.smegmatis* was isolated (see Materials and Methods); KK4.1 (see **Fig.4.1**) was used as primer and primer extension was performed as described in Materials and Methods. The result, illustrated in **Fig.7.6**, shows two products, with their 5'-ends both mapping to the stem region (the cutting sites are indicated in **Fig.7.1**). No specific product was seen in the reaction which contained no rRNA substrate. The majority of pre-rRNA was processed at a position identical to that detected in *B.subtilis* (Ogasawara et al., 1983). Because the first 20 nucleotides from the ITS region are identical to those of slow-growers, they also have the potential to pair with 5'- and 3'-ends of the 16S rRNA sequence. That the sequences are identical implies that the mechanisms of maturation of the 5'- and 3'-ends of the 16S

(A)

(a) GCCGGGGTGC (b) GCCGGGGTGC
(c) CGTTGAGATGCGGTGTG
GGAAACGCTGTTTCGATG

(B)

(a)

CGGG
| T
C-----G
5'G-----C3'

(b)

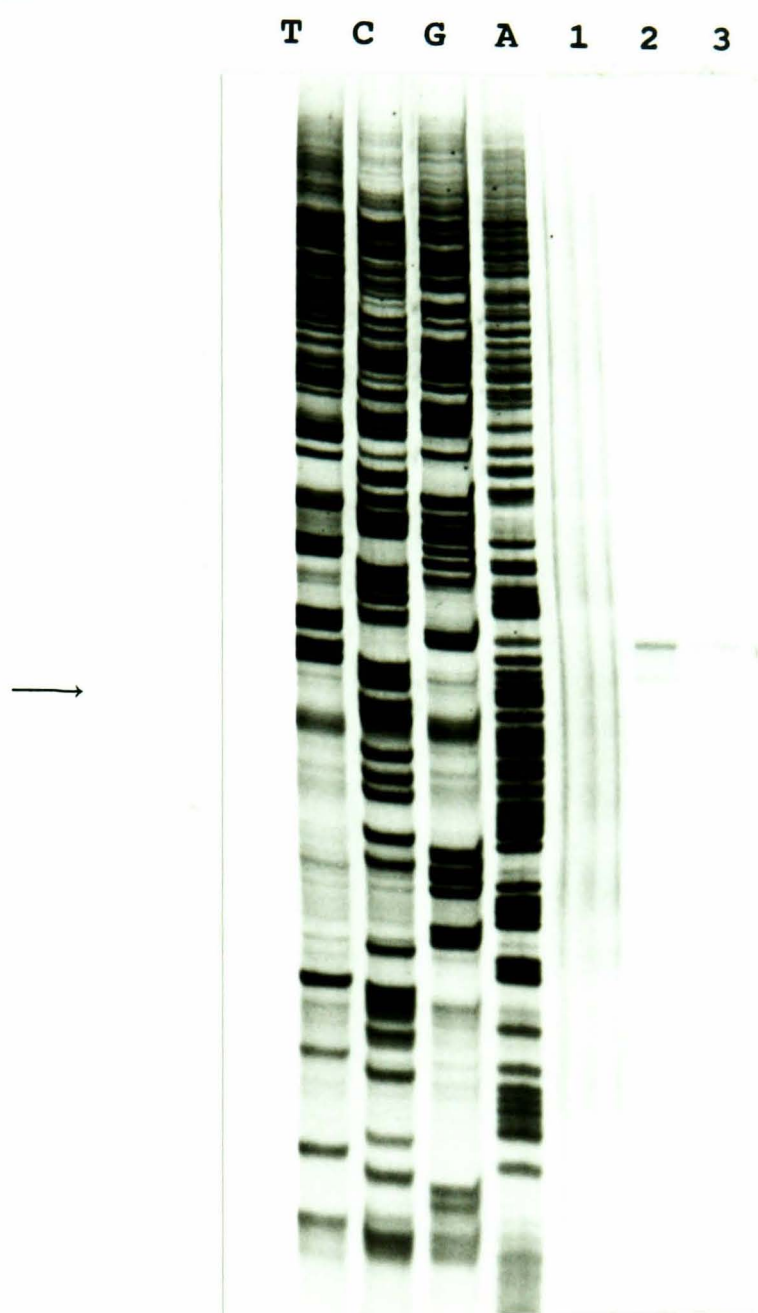
CGGG
| T
C-----G
5'G-----C3'

(c)

TGGG
G A
T-----A
G A
G-----C
C-----G
G-----C
T
T-----G
A-----T
G-----T
A-----T
G-----C
T-----G
T-----A
G-----T
5'C-----G3'

Figure 7.5 Comparison of helix S4 of *M.smegmatis* with that of slow-growers. The different helix S4 secondary structures within the ITS regions of slow-growing mycobacteria and *M.smegmatis*. (A) Shows the primary sequences of helix S4. (B) Shows the structures of a possible hairpin loop. (a) sequence from *M.leprae*; (b) sequence from *M.tuberculosis*; (c) sequence from *M.smegmatis*.

Fig.7.6 RNaseIII processing sites of *M.smegmatis* pre-16S rRNA



Identification of *in vivo* RNase III processing sites from *M.smegmatis* *rrn* operons by primer extension analysis. Lanes 1-3 represent primer KK4.1 (see Fig.4.1) - extension reactions, done with labelled KK.4.1 alone without RNA substrate, 24 μ g of isolated *M.smegmatis* NCTC 8159 RNA, and 12 μ g of isolated *M.smegmatis* NCTC 8159 RNA. Lanes **T**, **C**, **G** and **A** represent the sequencing reaction of the cloned upstream region of the 16S rRNA gene of *M.smegmatis* using primer **KK4.1**. The arrow indicates the position at which the majority of pre-rRNA was processed.

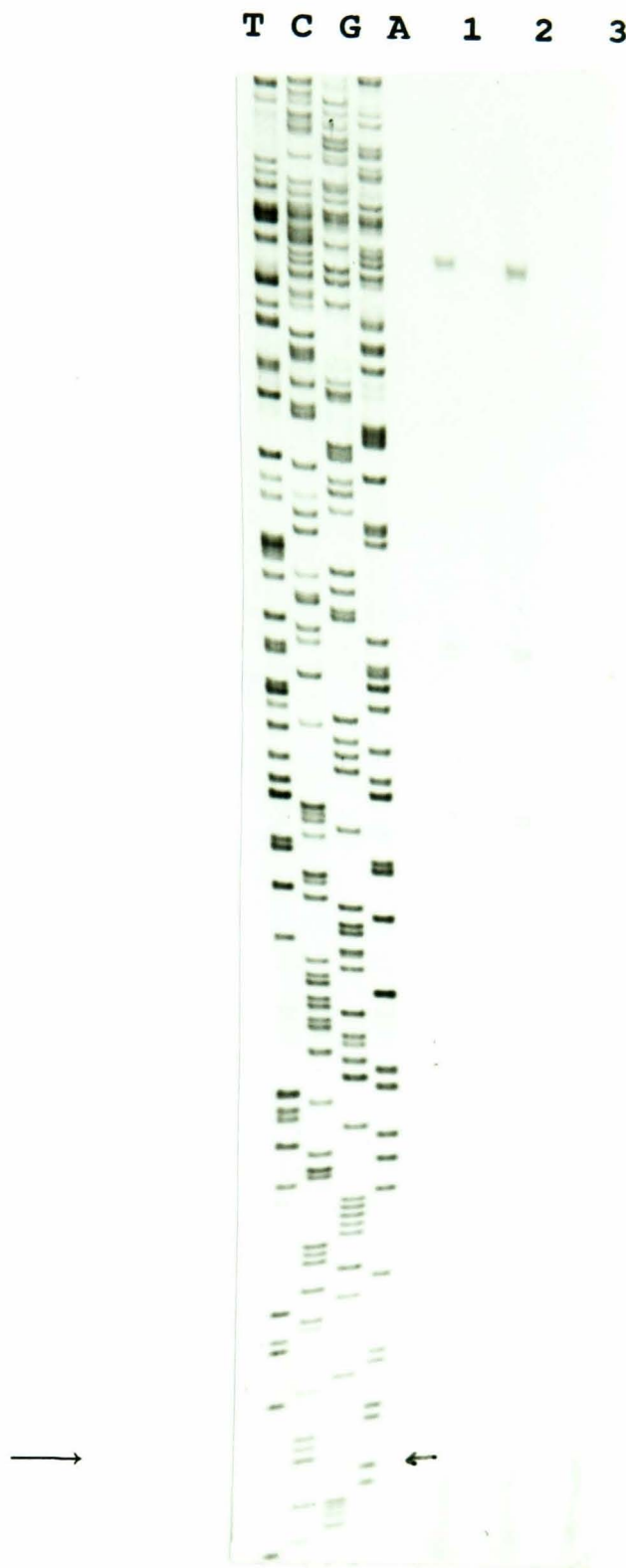
rRNA are similar to those of slow-growers (see Chapter 6).

7.2.6) Transcription start sites of the *M.smegmatis* *rrn* operons

The primer extension experiments in Chapter 3 showed that there is only one promoter in *rrn* operons from slow-growing mycobacteria and that the transcription start site is located within the conserved Box B element. The results in section 7.2.1 and 7.2.2 showed there are two different *rrn* operons in *M.smegmatis*. Both operons are identical in sequence downstream from the start of the Box A antitermination element. The *rrnI* operon has the Box B element, but the *rrnII* operon does not. This difference may be reflected in the transcription of *rrn* operons from fast-growing mycobacteria. It was interesting, therefore, to investigate the transcription pattern of such *rrn* operons, and to compare it with the pattern in slow-growing mycobacteria. Total RNA was isolated from *M.smegmatis* (see Materials and Methods); primer P-5 (see Fig.3.6) was end labelled with ^{32}P , and a primer extension experiment performed according to Materials and Methods.

Fig.7.7 shows the presence of three transcripts further upstream of the Box A element. Unlike the transcription start site of slow-growing mycobacteria, which was always detected at the Box B element, there was no transcript of the *M.smegmatis* *rrn* operon detected in the same position. The three transcripts were 72, 98 and 160⁺ bp upstream from the conserved Box B element, indicating

Fig.7.7 Transcription start sites of *M.smegmatis* *rrn* operons



Identification of the transcription start sites of the *rrn* operons of *M.smegmatis* NCTC 8159. Lanes 1-3 represent primer **P-5** extension experiment, done with 12 μ g RNA isolated from *M.smegmatis* cells, 24 μ g RNA isolated from *M.smegmatis* cells, and labelled primer **P-5** alone without RNA substrate, respectively. Lanes **T**, **C**, **G** and **A** represent sequencing reactions of the cloned upstream of the 16S rRNA of *M.smegmatis* using primer **P-5**. The arrow indicates the position where the transcription of the *rrn* operon of slow-growers starts.

that there were at least three promoters existing in the two *rrn* operons of *M.smegmatis*. Their positions were further upstream from the 5'-end of the 16S rRNA genes than that of the single promoter identified in *rrn* operons from slow-growing mycobacteria (see section 3.2.4).

7.3) Discussion

This Chapter describes how part of the 16S rRNA genes and their upstream sequences were cloned from both *rrn* operons of *M.smegmatis*. Both operons are identical in sequence in the region approximately 170 bp upstream from the 5'-end of the 16S rRNA gene. This region includes the putative Box A and Box C elements, and the RNase III processing sites. More interesting was the fact that these elements were also identical to those found in the slow-growing mycobacteria, providing further support for the notion that these elements play an important role in mycobacterial rRNA transcription and processing. The lack of a conserved Box B element in one of the two operons suggests that the two operons have different control mechanisms.

The primer extension experiment (**Fig.7.7**) suggested that at least three promoters were present in the two *rrn* operons, spanning 370 bp upstream from the 16S rRNA coding region. The two smaller transcripts were detected in previous experiments too; for example, they correspond to the two most intense bands in Fig.3.8. The consistent presence of these two bands further suggests that they are

genuine transcripts. The three promoters of the *M.smegmatis* *rrn* operons lie further upstream from the Box A sequence (or the RNase III processing site) than is the case for the single *rrn* operon of slow-growing mycobacteria. The experiment showed that *M.smegmatis* resembles *B.subtilis* and *E.coli*, where transcription of at least one *rrn* operon starts from multiple promoters, in contrast to the single promoter found in the *M.leprae* and *M.tuberculosis* *rrn* operons.

Although a longer ITS region was identified in *M.smegmatis* (354 bp compared with 280 bp in slow-growing mycobacteria), no tRNA gene could be found. The extra nucleotides were found to form a longer Helix S4 (see **Fig.7.5**). The significance of this observation is not clear, but a longer ITS region was also found in another fast-growing mycobacterium *M.phlei* (see **Fig.5.1**). Thus a longer Helix S4 may be characteristic of fast-growing mycobacteria, providing a criterion for the differentiation of slow and fast-growing mycobacteria.

Fig.7.6 shows the processing sites of *M.smegmatis* pre-16S rRNA at positions where more than 20 nucleotides are conserved in both fast- and slow-growing mycobacteria. This result suggests that a similar processing pathway exists in both fast and slow growers. A similar tract of sequences was found in *B.subtilis* (Ogasawara et al., 1983; and **Table 4.3**). This similarity suggests that the RNase III-like nuclease of mycobacteria may be sequence-specific, in the way that the corresponding *B.subtilis* enzyme appears to be.

This supports the suggestion that the mycobacterial rRNA genes are organized in an operon structure which is transcribed to form a precursor later to be processed into the different mature rRNA molecules.

CHAPTER EIGHT

General Discussion

The sequences that have been determined are rich in regions of significant interest, and are now available for the first time. They provide a description of the promoter regions of both *M.leprae* and *M.tuberculosis* *rrn* operons, the leader regions of *rrn* operons from eight species of slow-growing mycobacteria, and the intervening sequences between the 16S and 23S rRNA genes in the *rrn* operons. The data also show the sequences preceding both 16S rRNA genes of the fast-growing mycobacterium, *M.smegmatis*, together with the sequence from the intervening region between the 16S and 23S rRNA genes of *M.smegmatis*. These data make it possible to deduce important features of the organization and control of expression of rRNA genes in mycobacteria.

8.1) The organization of the mycobacterial *rrn* operon(s)

I have analyzed the sequences preceding the 16S rRNA gene and the spacer region of the 16S - 23S rRNA genes of mycobacteria. Using primer extension experiments in combination with DNA sequencing methods, I was able to define transcription start sites and processing sites.

By comparing the mode of rRNA synthesis in mycobacteria with that in other eubacteria such as *E.coli* and *B.subtilis*, several common features have emerged. Firstly, the 16S and 23S rRNA genes are arranged in

typical order, separated by 280 nucleotides in slow-growing mycobacteria and by 365 nucleotides in the fast-growing mycobacterium *M.smegmatis*. This spacer region bears no tRNA gene and therefore resembles eight out of ten of the *rrn* operons in *B.subtilis* (Loughney et al., 1982) and both *rrn* operons of *M.capricolum* (Sawada et al., 1984). Secondly, the existence of a primary transcript consisting of both 16S and 23S rRNA is indicated by the presence of promoter structures found only upstream of the 16S rRNA gene and not between the 16S rRNA gene and the 23S rRNA gene, by the repeated processing signals present, and by the observed precursor rRNAs. Thirdly, as in *E.coli* and *B.subtilis*, the primary transcript is first cleaved to yield premature 16S rRNA and perhaps precursor 23S rRNA as well; a consideration not included in this study. The cleavage sites fall within the putative stem-structures surrounding the sequences of the mature rRNAs (see **Fig.7.6**). These recognition sequences are homologous to corresponding regions of the rRNA precursors of *B.subtilis* and *Streptomyces*. They have also been discovered in the *rrnB* operon of *M.capricolum* and *M.hyopneumoniae* (see **Table 4.3**), but not in *E.coli* rRNA operons (Brosius et al., 1981).

In contrast to *rrn* operons of *E.coli* and some of the *B.subtilis*' *rrn* operons (Jinks-Robertson & Nomura, 1987; Ogasawara et al., 1983) transcription in *M.leprae* and *M.tuberculosis* (and perhaps other slow-growing mycobacteria) is initiated from a single promoter. This

arrangement resembles the organization found upstream from one of the *B.subtilis* 16S rRNA genes, where only one putative promoter has been found (Wawrousek & Hansen, 1983). However, it has not been established whether this promoter of *B.subtilis* is functional. As indicated by these authors, the expression of this 16S rRNA gene could be the result of transcriptional read-through from an upstream promoter. The *in vivo* primer extension experiments (Fig.3.7, 3.8 and Sela & Clark-Curtiss, 1991) prove that the identified promoters of *rrn* operons of *M.tuberculosis* and *M.leprae* are functional.

The three pre-rRNA transcripts identified from the total *M.smegmatis* RNA means that more than one promoter exists in one of the two *rrn* operons. The significance of such an organizational difference between the slower-growers and fast-growers is not clear. The results in Chapter 7 showed that the *rrnI* operon of *M.smegmatis* is slow-grower-like because it contains a conserved Box B element that has been found in all the slow-grower-species studied. The fact that, unlike the situation in slow-growing mycobacteria, no transcription start site was detected around the Box B element in the *M.smegmatis rrnI* operon, suggests that another control mechanism of rRNA synthesis is operating which is different from the one that exists in the slow-growing mycobacteria. Moreover, from the evolutionary point of view, the *rrnII* operon is likely to be the result of gene duplication.

In *E.coli*, the tandem promoters of *rrn* operons are

differentially regulated, the first (P1) being activated by ppGpp and responsible for the stringent control, whereas the second (P2) guarantees a basal level of expression (Glaser et al., 1983; Nomura et al., 1984). Evidence of growth-rate-dependent control of ribosome biosynthesis has been found in *M.tuberculosis* (Winder & Rooney, 1970). Since the equivalent of the structural P1 promoter is lacking in slow-growing mycobacteria, it would be interesting to know how the single promoter of this *rrn* operon fulfils the functions of two promoters in other bacteria. So far no significant homologies between upstream sequences preceding the promoters for 16S rRNA of *M.tuberculosis* and *M.leprae* have been detected. Development of a cloning system for slow-growing mycobacteria is a prerequisite for identifying the target sequences for these crucial functions.

8.2) The presence of antitermination elements in pre-16S rRNA of mycobacteria

A striking feature of eubacterial *rrn* operons is the presence of antitermination signals required to ensure complete transcription of the rRNA genes (Li et al., 1984; Morgan, 1986; Berg et al., 1989). Such signals are thought to be necessary to avoid problems arising from the absence of coupling between transcription and translation, or from premature transcription termination due to extensive secondary structure. In *E.coli* *rrn* operons, the antitermination site (Box A, 5'-TGCTCTTT-3') is located about 70 to 80 nucleotides upstream from the start of the

16S rRNA gene (Berg et al., 1989). These authors noted that antitermination-like features present in both the leader and the 16S-23S rRNA internal transcribed spacer region are remarkably conserved over a wide range of phylogenetically divergent micro-organisms, from eubacteria to archaeobacteria. In addition, positions 2 and 7 of Box A are invariant with position 3 having a high degree of variability. New experiments from Nodwell and Greenblatt (1993) showed directly that Box A RNA is a binding site of NusB and ribosome protein S10 and further suggested how Box A may act in subtly different ways in the phage and bacterial antitermination systems. NusB and S10 together (but neither alone) bind the *rrn* operon Box A, according to band shift analysis (Nodwell & Greenblatt, 1993). Furthermore, mutations that prevent antitermination activity *in vivo* of the ribosomal Box A also impair binding by NusB-S10. The Box A motif is located about 170 to 180 nucleotides upstream from the 16S rRNA gene of mycobacterial *rrn* operon(s). The invariant nucleotides at positions 2 and 7 are found in both the mycobacterial leader and in the ITS regions (see **Fig.4.2** and **Fig.5.2**). The main difference between the two mycobacterial Box A regions and those from *E.coli* is the occurrence of a G residue at the positions 3 and 5, which are Cs in *E.coli*. In contrast to *E.coli*, where the Box A region is partially base-paired, the Box A element in the leader region of mycobacterial *rrn* operon(s) could fully participate in secondary interactions within the proposed pre-16S rRNA

processing stem (see **Fig.6.1** and **Fig.6.2**). The same is true for other Gram-positive micro-organisms like *B.subtilis* and *M.capricolum* or *Streptomyces* (see **Table 4.3**). The fact that Box A is contiguous with the primary processing sites (see **Fig.6.1**, **Fig.6.2** and **Fig.7.6**) is intriguing and suggests a possible coupling between antitermination and processing.

Box B, a sequence of dyad symmetry and located 5' to Box A, is present in all seven *E.coli rrn* operons (Berg et al., 1989). Although Box A is capable of antitermination activity in the absence of Box B, point mutations have been found in Box B that decrease, or eliminate, terminator read-through in the presence of Box A. These mutations occur in the stem region of a putative hairpin structure, again suggesting that secondary structure might play a role in the antitermination process (Berg et al., 1989). The start site of transcription of the 16S rRNA gene from slow-growing mycobacteria is followed by a hairpin structure, including an internal loop that is reminiscent of a Box B antitermination motif. Although the consensus Box B sequences can not be identified in the ITS region of mycobacteria, a similar hairpin structure can still be formed by the sequence upstream of the Box A motif (see **Fig.5.2** and **Fig.7.4**). The high sequence divergence of Box B elements present in the leader and ITS regions of mycobacterial *rrn* operons confirms that the presence of dyad symmetry, and formation of potential hairpin structure, are fundamental for Box B function in antitermination.

Box C is an alternating GT sequence located after the Box B in Lambdoid nut sites and after Box A in the *rrn* leader antiterminator (Berg, et al., 1989). The function of Box C is not yet clear since antitermination negative mutations have not been found in any Box C region. Significant homology is found between the Box C region of *E.coli rrn* promoter-leader sequences and the equivalent region of the mycobacterial 16S rRNA promoter-leader sequence (Fig.4.2 and Fig.7.1). Although there is no Box C-like sequence following the *rrn* spacer Box A sequence in *E.coli*, Box C-like sequences seem to exist in the ITS region of mycobacteria (Fig.5.2 and Fig.7.4).

The high degree of similarity in the Box A sequence and the existence of dyad symmetry in the Box B element suggest that the principal features of the leader and ITS regions of the *E.coli rrn* operons are present, not only in the equivalent places of slow-growing mycobacteria, but also in fast-growing mycobacteria such as *M.smegmatis*.

8.3) Differentiation of *Mycobacterium* species by direct sequencing of the amplified leader or ITS region

rRNAs, in particular 16S rRNA, and their corresponding genes, are probably the molecules most frequently used at present for inferring evolutionary relationships among species on a molecular basis because of their high information content, conservative nature, and universal distribution (Fox et al., 1980; Woese, 1987). The 16S rRNA gene has been used frequently to distinguish different

mycobacterial species by means of PCR (Cox et al., 1991) or by direct sequencing after PCR (Rogall et al., 1990a; Rogall et al., 1990b). The greater than 94.3% sequence similarity of the 16S rRNA gene within the genus *Mycobacterium* (Rogall et al., 1990b) means that there are fewer than 75 nucleotide differences existing throughout the whole 16S rRNA gene. There are cases where the 16S rRNA sequences of closely related species are almost identical so these sequences are not sufficient for phylogenetic analysis (Fox et al., 1992). For example, the 16S rRNA sequences of *M. paratuberculosis* and *M. avium* are 99.9% identical (Rogall et al., 1990b). Although some more variable regions have been identified, for example helix 10 of the V2 region shown in Table 4.1, the existence of only one or two mismatched basepairs between closely related species, such as *M. marinum* and *M. tuberculosis*, makes it difficult to decide whether the nucleotide differences are genuine since an error of about 1/400 occurs when *Taq* polymerase is used in PCR (Saiki et al., 1988). *M. avium* and *M. paratuberculosis* are so closely related that they can not be distinguished using this stretch of nucleotide sequence (see Table 4.1).

The experiments with the leader (Chapter 4) and ITS (Chapter 5) regions of *rrn* operons of slow-growing mycobacteria revealed that these two regions are suitable for phylogenetic analysis of the genus *Mycobacterium*. The existence of a conserved Box B element in the leader region, and the conserved sequences at the end of the 16S

rRNA gene and the beginning of the 23S rRNA gene, makes the two regions accessible by PCR. The relative short amplified fragments, about 200 bp long for the leader region and about 280 bp long for the ITS region of slow growers, make them easy to sequence. Moreover the leader and ITS regions appear to be more variable than the 16S rRNA region, or even helix 10 of the V2 region which is thought to be highly variable in eubacteria (Woese et al., 1983), i.e. in positions 200 - 240 according to the numbering system for *E.coli* (Brosius et al., 1978) (see **Table 4.2** and **5.1**).

M.intracellulare, *M.avium*, and *M.paratuberculosis* form a discrete cluster. In many characteristics, *M.paratuberculosis* resembles *M.avium* (Merkal, 1984), and a close genetic relationship between *M.avium* and *M.paratuberculosis* has been suspected previously (Grange, 1984). The phenotypic diversity of organisms in the *M.avium* complex was illustrated by the results of recent DNA studies which indicated that *M.paratuberculosis* may be synonymous with *M.avium* (McFadden et al., 1987; Sommers & Good, 1985). The sequence analysis confirmed and reinforced this notion by finding a 100% similarity in the V2 regions of 16S rRNA (see **Table 4.1**). However, by comparing the leader regions of *M.avium* and *M.paratuberculosis* *rrn* operons, the two species can be defined by finding four nucleotide differences. This difference still places *M.paratuberculosis* as the most closely related species to *M.avium*, and proves that the leader region is commensurate with the spectrum of evolutionary distances inferred from

sequence studies on the 16S rRNA gene. The 16S rRNA sequences of *M.tuberculosis* and *M.marinum* are almost identical. However the eighty-eight nucleotide differences in the leader region suggest that the phylogenetic relationship of these species is more distant than the position established according to 16S rRNA gene sequences (Rogall et al., 1990b).

The results presented in Chapter 4 and Chapter 5 reveal that the leader and ITS regions of mycobacteria are readily amplified by PCR. Thus the nucleotide sequences of these regions provide additional information for establishing phylogenetic relationships. The greater variation in sequence found between species for these two regions, compared with the highly conserved 16S rRNA sequences, provides a more reliable way of discriminating between very closely related species. Further work to investigate the two regions from more *Mycobacterium* species needs to be undertaken. Once completed, the available sequencing data can be used as another criterion for the classification of species in the genus *Mycobacterium*.

8.4) Expression of foreign antigens

One of the aims of work on the mycobacterial *rrn* operon was to explore the possibility of using the promoter to express other mycobacterial genes either in *E.coli* or in a fast-growing mycobacterium. A strong promoter would be ideal for expressing massive quantities of recombinant gene products. The promoter for the *rrn* operon might fulfil this

criterion, since this promoter can work in *E.coli* as well as in mycobacteria. Gene products could be expected to be produced in quantity in either system. If expression of foreign genes can be achieved in *E.coli*, then this promoter will provide an excellent basis for the development of vectors for the expression of mycobacterial genes. Both fragments containing the *M.leprae* and *M.tuberculosis rrn* promoters were successfully used to drive expression of the CAT gene; however their ability to express other mycobacterial genes has not been tested. The promoter could be used for the expression of such genes as that for the *M.leprae* 65 kDa heat shock protein (Mehra et al., 1986), for which species-specific monoclonal antibodies are available. Alternatively, the gene encoding the 19 kDa antigen of *M.tuberculosis* (Ashbridge et al., 1989) and *M.bovis* BCG (Collins et al., 1990), for which a specific monoclonal antibody exists, could be expressed.

8.5) Future work

8.5.1) Cloning the promoters of the two *rrn* operons of *M.smegmatis*

This area of work now requires investigation. Identification of the promoter sequences would provide further elucidation of the transcription start sites and may shed some light on the control mechanism for rRNA synthesis. Several methods are commonly used for cloning an unknown region, including inverse PCR and screening a mini-library.

Inverse PCR has been used in Chapter 7 to clone the upstream sequence of the 16S rRNA gene of the *rrnII* operon. Because the 1.0 kbp amplified fragment is too short to contain any sequence from the promoter region, the genomic DNA digested with *Pst*I enzyme can be used as the PCR substrate. If the same pair of primers illustrated in Fig.7.3 were used, two products with sizes of 1.2 kbp and 3.0 kbp each would be expected according to the restriction maps of the two *rrn* operons (Bercovier et al., 1989). Once the two fragments are sequenced, together with the help of the transcription start sites identified in Fig.7.7, the promoter sequence at the -10 and -35 regions could be correctly determined and analyzed.

The other possible method is by cloning DNA fragments in the range 3.2 and 1.4 kbp after *Pst*I enzyme digestion. The two cloning pools could then be used to transform *E.coli* and the colonies screened by a riboprobe. The riboprobe could be made as follows: the DNA containing the leader region of either of the two operons (see Fig.7.1) is first cloned into a vector containing a promoter recognized by a bacteriophage RNA polymerase. The DNA is cloned in the orientation which will result in the production of antisense RNA. The highly labelled antisense rRNA transcripts thus produced from the cloned DNA can then be used as probes to hybridize the *E.coli* colonies which contain the clones of interest. After screening, positive colonies can then be sequenced.

8.5.2) Determination of the number of pre-rRNA transcripts from each of the *M.smegmatis* rrn operons

The exact number of pre-rRNA transcripts from each of the rrn operons can be determined by primer extension. The upstream sequences of the two 16S rRNA genes are different up from the Box A element (see Fig.7.1). These different sequences could be used as targets for specific primers. For example, the sequence complementary to the Box B element can be used as one operon specific primer. This primer then is only able to bind to the rrnI, and the primer extension products are sure to originate from rrnI but not from rrnII. Such approaches will help to ascertain if the tandem promoters exist in one rrn operon or in both.

8.5.3) Establishing sequences of the leader and ITS regions from more mycobacterial species

The leader and ITS sequences established from eight and six mycobacterial species, respectively, shed light on the value of these two areas for the investigation of phylogenetic relationships among species of *Mycobacterium* genus. To make the approach practical, more species need to be studied. The methods for establishing the leader and ITS sequences are given in Chapter 4 and Chapter 5. Amplifying the ITS region can be achieved by using primer combination JCA 3/KK 2 instead of JCA 3/RAC 5.1, the combination used in Chapter 5. This difference will make the amplified fragment shorter and easy to clone and sequence.

8.5.4) Identification of individual mycobacteria

To take advantage of the abundance of rRNA in cells,

a new method using fluorescence microscopy has been developed to identify single cells by labelled fluorescence tagged oligodeoxynucleotides hybridized to their 16S rRNA (DeLong et al., 1989). Labelled oligo-probes were found to enter the target cells freely without pre-treatment of the cells. It is proposed that this method be extended to the identification of slow and fast-growing mycobacteria. It is possible to use the fluorescently-labelled probes to detect individual mycobacteria. A number of probes complementary to the 16S rRNA are specific for fast and slow-growing mycobacteria. One which is specific for *M.leprae* has been developed in this laboratory (Cox et al., 1991). This approach makes it likely that fluorescently labelled probes specific for *M.leprae* could be used to diagnose the presence of *M.leprae* in the tissue of leprosy patients. In addition the fluorescence intensity of a single cell due to hybridization with the probe varies linearly with the growth rate and parallels measured and known values for RNA per cell, and for ratios of RNA to DNA content (DeLong et al., 1989). The slow growth rate of pathogenic mycobacteria and our inability to cultivate *M.leprae* make it difficult to measure growth rate. Measurements of fluorescence intensity of the stained mycobacteria, for example *M.leprae*, should reflect growth rate. Development of such an *in situ* hybridization system in mycobacteria would benefit early diagnosis of diseases of mycobacterial origin, and analysis of the effectiveness of new anti *Mycobacterium* drugs.

8.6) Conclusion

The results presented in this thesis demonstrate that the promoter and leader regions of mycobacterial *rrn* operon(s) are similar to those of other bacteria. The promoter functioned reasonably well in *E.coli*, and the antitermination elements were identified in both leader and ITS regions. The sequences established for the leader and ITS regions from several species unveiled another two areas for phylogenetic analysis and species identification. It would be possible to use the techniques developed in this project to establish the leader and ITS sequences of more species, in order to define phylogenetic relationships among the *Mycobacterium* genus.

The mechanisms for controlling the transcription of the *rrn* operons of slow- and fast-growing mycobacteria differ in detail. For example, the *rrn* operon of slow-growers is controlled by a single promoter whereas the *rrn* operons of *M.smegmatis* are controlled by more than one promoter. Finally, details of the organisation of the three or more promoters of *M.smegmatis* need to be established. Further work should enable one to clone the promoters of the two *rrn* operons of *M.smegmatis*, and to identify precisely the number of promoters in each operon.

ACKNOWLEDGEMENTS

I should like to thank Dr. John Skehel and Dr. J.R. Tata for allowing me to carry out this research at the Department of Developmental Biochemistry in the National Institute for Medical Research.

It is a great pleasure to thank my supervisor Dr. Robert Cox for continual enthusiasm, guidance and encouragement throughout the course of my project. His treatment of my sometimes impatient character with great leniency makes me laugh at my naivety.

I am especially grateful to Dr. Jo Colston for providing me with the chance to work in the UK and his patience in reading through my Chinese-English thesis. His support and encouragement are always unlimited.

Thanks also go to Dr. Qiling Xu for her enormous help in teaching me basic molecular biology when I first arrived. She and Dr. Dave Wilkinson made our stay in the UK more colourful.

I thank Dr. Elaine Davis for her donation of plasmid vectors; Dr. Tanya Parish for providing some *M. smegmatis* DNA; Paul Thurman for growing some mycobacterial cultures. Finally, I wish to thank my parents who do not realize how important their contribution has been although they are thousands of miles away.

APPENDIX A: Media

Lemco Broth

Bacto-peptone	10 g
Bacto Lab Lemco powder	5 g
NaCl	5 g
Distilled water	to 1000 ml
pH to 7.2 with 2 M NaOH, sterilize by autoclaving	

Luria - Bertani Medium (LB)

Bacto-tryptone	10 g
Bacto-yeast extract	5 g
NaCl	10 g
Distilled water	to 1000 ml
sterilize by autoclaving	

Modified Dubos Medium

Part A

KH_2PO_4	1 g
$\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$	6.25 g
$\text{Na}_3\text{citrate}$	1.25 g
$\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$	0.6 g
Asparagine	2 g
10% Tween 80	5 ml
Casamino acids	2 g

Dissolve asparagine in 50 ml hot water, add the rest of the ingredients, make up to 950 ml, adjust pH to 7.2 with 2 M NaOH, make up to 1000 ml, sterilize by autoclaving

Part B

Bovine Serum Albumin Fraction V	5 g
Distilled water	to 1000 ml

filter sterilize

for use add 10 ml part B to 250 ml part A

2xTY Medium

Bacto-tryptone	16 g
Bacto-yeast extract	10 g
NaCl	5 g
Distilled water	to 1000 ml

Adjust pH to 7 with NaOH and sterilize by autoclaving

APPENDIX B: Recombinant plasmids

(A) pUC19 based recombinant plasmids, contain the specified inserts in pUC19:

KEK6	1.8 kbp <i>Pst</i> I fragment containing part of the <i>M.leprae</i> 16S rRNA gene and its upstream sequence
PIE1	1.2 kbp <i>Pst</i> I fragment of part of the <i>M.tuberculosis</i> 16S rRNA gene and its upstream sequence

(B) pBluescript based recombinant plasmids, contain the specified inserts in pBluescript:

KEK6.1	685 bp <i>Pst</i> I- <i>Bam</i> HI fragment of the <i>M.leprae</i> 16S rRNA gene upstream sequence
KEK6.2	1.115 kbp <i>Bam</i> HI- <i>Pst</i> I fragment of part of the <i>M.leprae</i> 16S rRNA gene and its upstream sequence
pJY105	Inverse PCR fragment of <i>M.smegmatis</i> <i>rrn</i> II containing 233 bp upstream of the 16S rRNA gene and 767 bp of the 16S rRNA gene

All the PCR fragments amplified by the primer combinations **PL1/P4** and **JCA3/RAC5.1** were cloned into the *Eco*RV restriction site of pBluescript

(C) pEJ106 based recombinant plasmids, contain the specified inserts in pEJ106:

pJY101 containing the same insert as KEK6; the insert is
 in the correct orientation to allow the
 expression of CAT gene

pJY102 as the plasmid pJY101 but the insert orientation
 is different

pJY104 containing the same insert as PIE1; the insert is
 in the correct orientation to allow the
 expression of CAT gene

pJY103 as the plasmid pJY104 but the insert orientation
 is different

REFERENCES

- Abou-Zeid,C., Smith,E., Grange,J.M., Ratliff,T.L., Steele,J. and Rook,G.A.W.** (1988): The secreted antigens of *Mycobacterium tuberculosis* and their relationship to those recognized by the available antibodies. *Journal of General Microbiology* **134**, 531-538.
- Aksoy,S., Squires,C.L. and Squires,C.** (1984): Evidence for antitermination in *Escherichia coli* rRNA transcription. *Journal of Bacteriology* **159** (1), 260-264.
- Altman,S.** (1975): Biosynthesis of transfer RNA in *Escherichia coli*. *Cell* **4**, 21-29.
- Apirion,D. and Miczak,A.** (1993): RNA processing in prokaryotic cells. *BioEssays* **15** (2), 113-120.
- Arnstein,H.R.V. and Cox,R.A.** (1992): Ribosome structure. In: *Protein biosynthesis*. (Eds: Rickwood,D; Male,D) Oxford University Press, New York. 35-55.
- Ashbridge,K.R., Booth,R.J., Watson,J.D. and Lathigra,R.B.** (1989): Nucleotide sequence of the 19kDa antigen gene from *Mycobacterium tuberculosis*. *Nucleic Acids Research* **17**, 1249.
- Azad,A.A.** (1979): Intermolecular base-paired interaction

between complementary sequences present near the 3'-end of 5S rRNA and 18S (16S) rRNA might be involved in the reversible association of ribosomal subunits. *Nucleic Acids Research* 7, 1913-1929.

Baess, I. (1979): Deoxyribonucleic acid relatedness among species of slowly growing mycobacteria. *Acta Pathologica, Microbiologica et Immunologica Scandinavica* B87, 221-226.

Baess, I. (1983): Deoxyribonucleic acid relationships between different serovars of *Mycobacterium avium*, *Mycobacterium intracellulare* and *Mycobacterium scrofulaceum*. *Acta Pathologica, Microbiologica et Immunologica Scandinavica* B91, 201-203.

Baird, P.N., Hall, L.M. and Coates, A.R. (1988): A major antigen from *Mycobacterium tuberculosis* which is homologous to the heat shock proteins groES from *E.coli* and the hptA gene product of *Coxiella burnetii*. *Nucleic Acids Research* 18, 9047.

Bercovier, H., Kafri, O., Kornitzer, D. and Sela, S. (1989): Cloning and restriction analysis of ribosomal RNA genes from *Mycobacterium smegmatis*. *FEMS Microbiology Letters* 57, 125-128.

Bercovier, H., Kafri, O. and Sela, S. (1986): *Mycobacteria* possess a surprisingly small number of ribosomal RNA genes

in relation to the size of their genome. *Biochemical and Biophysical Research Communications* **136**, 1136-1141.

Berg, K.L., Squires, C. and Squires, C.L. (1989): Ribosomal RNA operon anti-termination: function of leader and spacer region BoxB - BoxA A sequences and their conservation in diverse microorganisms. *Journal of Molecular Biology* **209**, 345-358.

Borremans, M., de Wit, L., Volckaert, G., Ooms, J., de Bruyn, J., Huygen, K., van Vooren, J.P., Stelande, M., Verhofstadt, R. and Content, J. (1989): Cloning, sequence determination, and expression of a 32-kilodalton-protein gene of *Mycobacterium tuberculosis*. *Infection and Immunity* **57**, 3123-3130.

Bradley, S.G. (1973): Relationships among mycobacteria and nocardiae based upon deoxyribonucleic acid reassociation. *Journal of Bacteriology* **113**, 645-651.

Bram, R.J., Young, R.A. and Steitz, J.A. (1980): The Ribonuclease III site flanking 23S sequence in the 30S ribosomal precursor RNA of *E.coli*. *Cell* **19**, 393-401.

Bremer, H. and Dennis, P.P. (1987): . In: *Escherichia coli and Salmonella typhimurium Cellular and Molecular Biology*. (Ed: Neidhardt, FC et al) American Society for Microbiology, Washington DC. 1527-1542.

Brewster, J.M. and Morgan, E.A. (1981): Tn9 and IS1 inserts in a ribosomal ribonucleic acid operon of *Escherichia coli* are incompletely polar. *Journal of Bacteriology* **148**, 897-903.

Brimacombe, R., Maly, P. and Zwieb, C. (1983): The structure of ribosomal RNA and its organization relative to ribosomal protein. *Progress in Nucleic Acid Research Molecular Biology* **28**, 1-48.

Brosius, J., Dull, T.J., Sleeter, D.D. and Noller, H.F. (1981): Gene organization and primary structure of a ribosomal RNA operon from *Escherichia coli*. *Journal of Molecular Biology* **148**, 107-127.

Brosius, J., Palmer, M.L., Kennedy, P.L. and Noller, H.F. (1978): Complete nucleotide sequence of a 16S ribosomal RNA gene from *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **75**, 4801-4805.

Chamberlain, M.J. (1982): Chapter 3 Bacterial DNA-dependent RNA polymerases. In: *The Enzymes*. Vol. 15. (Ed: Boyer, PD) Academic Press, London. 61-86.

Chiodini, R.J. (1989): Crohn's disease and the mycobacterioses: a review and comparison of two disease entities. *Clinical Microbiological Review* **2**, 90-117.

Clark, J.M. (1988): Novel non-templated nucleotide addition reactions catalyzed by prokaryotic and eukaryotic DNA polymerases. *Nucleic Acids Research* **16**, 9677-9686.

Clark-Curtiss, J.E. (1990): Genome structure of mycobacteria. In: *Molecular Biology of the Mycobacteria*. (Ed: McFadden, J) Surrey University Press, Surrey, U.K. 77-96.

Clark-Curtiss, J.E., Jacobs, R.W., Docherty, M.A., Richie, L.R. and Curtiss III, R.W. (1985): The molecular analysis of DNA and the construction of genomic libraries of *M. leprae*. *Journal of Bacteriology* **161**, 1093-1102.

Clarke, P.H. and Meadow, P.M. (1959): Evidence for the occurrence of permeases for tricarboxylic acid cycle intermediates in *Pseudomonas aeruginosa*. *Journal of General Microbiology* **20**, 144-155.

Cohen, S.N., Chang, A.C. and Hsu, L. (1972): Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *E. coli* by R-factor DNA. *Proceedings of the National Academy of Sciences of the United States of America* **69**, 2110-2114.

Collado-Vibes, J., Magasanik, B. and Gralla, J.D. (1991): Control site location and transcriptional regulation in *Escherichia coli*. *Microbiological Reviews* **55**, 371-394.

Collins,F.M. (1988): AIDS-related mycobacterial disease. *Springer Seminar of Immunopathology* **10**, 375-391.

Collins,M.E., Patki,A., Wall,S., Nolan,A., Goodger,J., Woodward,M.J. and Dale,J.W. (1990): Cloning and characterization of the gene for the 19kDa antigen of *Mycobacterium bovis*. *Journal of General Microbiology* **136**, 1429-1436.

Cox,R.A., Kempsell,K., Fairclough,L. and Clston,M.J. (1991): The 16S ribosomal RNA of *Mycobacterium leprae* contains a unique sequence which can be used for identification by the polymerase chain reaction. *Journal of Medical Microbiology* **35**, 284-290.

Crawford,J.T., Cave,D.M. and Bates,J.H. (1981): Evidence for plasmid mediated restriction modification in *mycobacterium avium intracellulare*. *Journal of General Microbiology* **127**, 333-338.

Cross,S.H. and Little,P.F. (1986): A cosmid vector for systematic chromosome walking. *Gene* **49**, 9-22.

Cruickshank,R. (1965): *Medical Microbiology: a guide to the laboratory diagnosis and control of infection*. 11th ed. E & S Livingstone Limited, London. Ch**44**, 652-655.

Csordás-Tóth,E., Boros,E. and Venelianer,P. (1979):

Structure of the promoter region for the *rrnB* gene in *Escherichia coli*. *Nucleic Acids Research* **7**, 2189-2197.

Dahlberg,A.E. (1989): The functional role of ribosomal RNA in protein synthesis. *Cell* **57**, 525-529.

Dahlberg,A.E., Dahlberg,J.E., Lund,E., Tokimatsu,H. and Rabson,A.B. (1978): Processing of the 5'-end of *Escherichia coli* 16S ribosomal RNA. *Proceedings of the National Academy of Sciences of the United States of America* **75**, 3598-3602.

Dale,J.W. and Patki,A. (1990): Mycobacterial gene expression and regulation. In *Molecular Biology of the Mycobacteria*, pp. 173-198. Edited by J.McFadden. London: Surrey University Press.

Dams,E., Hendriks,L., Van de Peer,Y., Neefs,J.M., Smits,G., Vandenkempt,J. and de Wachter,R. (1988): Compilation of small ribosomal subunit RNA sequences. *Nucleic Acids Research* **16**, r87-r175.

Daniels,D.L., Subbarao,M.N., Blattner,F.R. and Lozeron,H.A. (1988): Q-Mediated late gene transcription of bacteriophage λ : RNA start point and RNase III processing sites *in vivo*. *Virology* **167**, 568-577.

deBoer,H.A., Gilbert,S.F. and Nomura,M. (1979): DNA sequences of promoter regions for rRNA operons *rrnE* and

rrnA in *E.coli*. *Cell* **17**, 201-209.

DeLong, E.F., Wickham, G.S. and Pace, N.R. (1989): Phylogenetic stains: ribosomal RNA-based probes for the identification of single cells. *Science* **243**, 1360-1363.

Deutscher, M.P. (1984): Processing of tRNA in prokaryotes and eukaryotes. *CRC Critical Reviews in Biochemistry* **17**, 45.

Dubos, R.J. and Davis, B.D. (1946): Factors affecting the growth of tubercle bacilli in liquid media. *Journal of Experimental Medicine* **83**, 409-423.

Duester, G.L. and Holmes, W.M. (1980): The distal end of the ribosomal RNA operon *rrnD* of *Escherichia coli* contains a tRNA gene, two 5S rRNA genes and a transcription terminator. *Nucleic Acids Research* **8**, 3793-3806.

Ellis, J. (1987): Proteins as molecular chaperones. *Nature* **328**, 378-379.

Elsaghier, A., Prantera, C., Moreno, C. and Ivanyi, J. (1992): Antibodies to *Mycobacterium paratuberculosis*-specific protein antigens in Crohn's disease. *Clinical and Experimental Immunology* **90**, 503-508.

Eperon, P.L., Estibeiro, J.P. and Eperon, I.C. (1986): The role of nucleotide sequences in splice site selection in

eukaryotic pre-messenger RNA. *Nature* **324**, 280-282.

Feunteun, J., Jordan, B.R. and Monier, R. (1972): Study of the maturation of 5S RNA precursors in *Escherichia coli*. *Journal of Molecular Biology* **70**, 465-474.

Fine, P.E.M. and Rodrigues, L.C. (1990): Modern vaccine. Mycobacterial diseases. *Lancet* **335**, 1016-1020.

Fite, G.L., Cambre, P.J. and Turner, M.H. (1947): Procedure for demonstrating lepra bacilli in paraffin sections. *Archives of Pathology* **43**, 624-627.

Fox, G.E., Stackebrandt, E., Hespell, R.B., Gibson, J., Maniloff, J., Dyer, T.A., Wolfe, R.S., Balch, W.E., Tanner, R.S., Magrum, L.J., Zablen, L.B., Blackmore, R., Gupta, R., Bonen, L., Lewis, B.J., Stahl, D.A., Luehrsen, K.R., Chen, K.N. and Woese, C.R. (1980): The phylogeny of prokaryotes. *Science* **209**, 457-463.

Friedman, D.I. and Gottesman, M. (1983): The lytic mode of lambda development. In: *Lambda II*. (Eds: Hendrix, R; Roberts, J; Stahl, F; Weisberg, R) Cold Spring Harbor laboratory, Cold Spring Harbor, New York. 21-52.

Gallant, J. (1979): Stringent control in *E.coli*. *Annual Review of Genetics* **13**, 393-415.

Garsia,R.J., Hellqvist,L., Booth,R.J., Radford,A.J., Britton,W.J., Astbury,L., Trent,R.J. and Basten,A. (1989): Homology of the 70-kilodalton antigens from *Mycobacterium leprae* and *Mycobacterium tuberculosis* 71-kilodalton antigen and with the conserved heat shock protein 70 of eucaryotes. *Infection and Immunity* **57**, 204-212.

Gartenberg,M.R. and Crothers,D.M. (1988): DNA sequence determinants of CAP-induced bending and protein binding affinity. *Nature* **333**, 824-826.

Gegenheimer,P. and Apirion,D. (1981): Processing of procaryotic ribonucleic acid. *Microbiological Reviews* **45**, 502-541.

Gegenheimer,P., Watson,N. and Apirion,D. (1977): Multiple pathways for primary processing of ribosomal RNA in *Escherichia coli*. *Journal of Biological Chemistry* **252**, 3064-3073.

Girvitz,S.C., Bachetti,S., Rainbow,A.J. and Graham,F.L. (1980): A rapid and efficient procedure for the purification of DNA from agarose gels. *Analytical Biochemistry* **106**, 492-496.

Gitt,M.A., Wang,L.F. and Doi,R.H. (1985): A strong sequence homology exists between the major RNA polymerase sigma factors of *Bacillus subtilis* and *Escherichia coli*. *The*

Journal of Biological Chemistry **260**, 7178-7185.

Glaser, G. and Cashel, M. (1979): In vitro transcription from the *rrn B* ribosomal RNA cistron originates from two tandem promoters. *Cell* **16**, 111-121.

Glaser, G., Sarmientos, P. and Cashel, M. (1983): Functional interrelationship between two tandem *E.coli* ribosomal RNA promoters. *Nature* **302**, 74-76.

Good, R.C. (1985): Opportunistic pathogens in the genus *Mycobacterium*. *Annual Review of Microbiology* **39**, 347-369.

Goodfellow, M. and Wayne, L.G. (1982): Taxonomy and nomenclature. In: *The Biology of the Mycobacteria*. Vol. 1. (Eds: Ratledge, C; Stanford, J) Academic Press, London, Chapter **10**, 471-521.

Gourse, R., deBoer, H.A. and Nomura, M. (1986): DNA determinants of rRNA synthesis in *E.coli*: Growth rate dependent regulation, feedback inhibition, upstream activation, antitermination. *Cell* **44**, 197-205.

Gralla, J.D. (1990): Promoter recognition and mRNA initiation by *Escherichia coli* E-sigma-70. *Methods in Enzymology* **185**, 37-54.

Grange, J.M. (1982): The genetics of *Mycobacteria* and

Mycobacteriophages. In: *The Biology of the Mycobacteria*. Vol. 1. (Eds: Ratledge,C; Stanford,J) Academic Press.

Grange,J.M. (1984): *Mycobacterium avium*. *European Journal of Respiratory Disease* **65**, 399-401.

Gray,M.W., Sankoff,D. and Cedargreen,R.J. (1984): On the evolutionary descent of organisms and organelles; a global phylogeny based on a highly conserved structure core in small subunit ribosomal RNA. *Nucleic Acids Research* **12**, 5837-5852.

Gross,W.M. and Wayne,L.G. (1970): Nucleic acid homology within the genus *Mycobacterium*. *Journal of Bacteriology* **104**, 630-634.

Hansen,G.A. (1874): Undersøgelser angående spedalskhedens årsager. *Norsk Magazin for Laegevitenskapen* **4**, 1-88.

Harley,C.B. and Reynolds,R.P. (1987): Analysis of *E.coli* promoter sequences. *Nucleic Acids Research* **15**, 2343-2361.

Harvey,S., Hill,C.W., Squires,C. and Squires,C.L. (1988): Loss of the spacer loop sequence from the *rrnB* operon in the *Escherichia coli* K-12 subline that bears the *relA1* mutation. *Journal of Bacteriology* **170**(3), 1235-1238.

Helmann,J.D. and Chamberlain,M.J. (1988): Structure and

function of bacterial sigma factors. *Annual Review of Biochemistry* **57**, 839-872.

Holben, W.E. and Morgan, E.A. (1984): Antitermination of transcription from an *Escherichia coli* ribosomal RNA promoter. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 6789-6793.

Holben, W.E., Prasad, S.M. and Morgan, E.A. (1985): Antitermination by both the promoter and the leader regions of an *Escherichia coli* ribosomal RNA operon. *Proceedings of the National Academy of Sciences of the United States of America* **82**, 5073-5077.

Hottat, F., Coene, M. and Cocito, C. (1988): DNA methylation in leprosy-associated bacteria: *Mycobacterium leprae* and *Corynebacterium tuberculo-stearicum*. *Medical Microbiology and Immunology* **177**, 33-45.

Husson, R.N., James, B.E. and Young, R.A. (1990): Gene replacement and expression of foreign DNA in *Mycobacteria*. *Journal of Bacteriology* **172**(2), 519-524.

Imaeda, T., Broslawski, G. and Imaeda, S. (1988): Genomic relatedness among mycobacterial species by nonisotopic blot hybridization. *International Journal of Systemic Bacteriology* **38**, 151-156.

Imaeda,T., Kirchheimer,W.F. and Barksdale,L. (1982): DNA isolated from *Mycobacterium leprae*: genome size, base ratio, and homology with other related bacteria as determined by optical DNA-DNA reassociation. *Journal of Bacteriology* **150**, 414-417.

Iwami,M., Muto,A., Yamao,F. and Osawa,S. (1984): Nucleotide sequence of the *rrnB* 16S ribosomal RNA gene from *Mycoplasma capricolum*. *Molecular and General Genetics* **196**, 317-322.

Jacobs,W.R., Tuckman,M. and Bloom,B.R. (1987): Introduction of foreign DNA into mycobacteria using a shuttle plasmid. *Nature* **327**, 532-535.

Jacobs Jr,W.R., Docherty,M.A., Curtiss III,R. and Clark-Curtiss,J.E. (1986): Expression of *Mycobacterium leprae* genes from a *Streptococcus* mutant promoter in *Escherichia coli* K-12. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 1926-1930.

Ji,Y.E., Colston,M.J. and Cox,R.A. (1994): Nucleotide sequence and secondary structures of precursor 16S rRNA of slow-growing mycobacteria. *Microbiology* **1**, in press.

Jinks-Robertson,S. and Nomura,M. (1987): *Cellular and Molecular Biology*. In: *Escherichia coli and Salmonella typhimurium*. Vol. 2. (Eds: Neidhardt,FC; Ingraham,JL; Brooks Low,K; Magasanik,B; Schaechter,M; Umberger,HE)

American Society for Microbiology, Washington, DC. 1358-1385.

Katoch,V. and Cox,R.A. (1986): Stepwise isolation of RNA and DNA from mycobacteria. *International Journal of Leprosy* **54**, 409-415.

Kempsell,K.E., Ji,Y., Estrada-G,I.C.E., Colston,M.J. and Cox,R.A. (1992): The nucleotide sequence of the promoter, 16S rRNA and spacer region of the ribosomal RNA operon of *Mycobacterium tuberculosis* and comparison with *Mycobacterium leprae* precursor rRNA. *Journal of General Microbiology* **138**, 1717-1727.

King,T.C. and Schlessinger,D. (1983): S1 nuclease mapping analysis of ribosomal RNA processing in wild-type and processing deficient *Escherichia coli*. *Journal of Biological Chemistry* **258**, 12034-12042.

Kinger,A.K., Verma,A. and Tyagi,J.S. (1993): A method for the isolation of pure intact RNA from mycobacteria. *Biotechniques* **14**, 724-725.

Koch,R. (1882): Die Ätiologie der Tuberkulose. *Berliner Klinischen Wochenschrift* **15**, 221-230.

Kochi,A. (1991): The global tuberculosis situation and the new control strategy of the World Health Organization.

Tubercle 72, 1-6.

Krinke, L. and Wulff, D.L. (1990): The cleavage specificity of RNase III. *Nucleic Acids Research* 18, 4809-4815.

Lamond, A.I. and Travers, A.A. (1985): Genetically separable functional elements mediate the optimal expression and stringent regulation of a bacterial gene. *Cell* 40, 319-326.

Lehmann, K.B. and Neumann, R. (1896): *Atlas und Grundriss der Bakteriologie und Lehrbuch der speciellen bakteriologischen Diagnostik*. 1st ed. Lehmann, J.F., Munchen, Germany.

Li, J., Horwitz, R., McCracken, S. and Greenblatt, J. (1992): NusG, a new *E.coli* elongation factor required for processive antitermination of transcription by the N protein of phage λ . *Journal of Biological Chemistry* 267, 6012-6019.

Li, S.C., Squires, C.L. and Squires, C. (1984): Antitermination of *E.coli* rRNA transcription is caused by a control region segment containing nut-like sequences. *Cell* 38, 851-860.

Liesack, W., Pitulle, C., Sela, S. and Stackebrandt, E. (1990): Nucleotide sequence of the 16S rRNA from *Mycobacterium leprae*. *Nucleic Acids Research* 18, 5558.

**Liesack,W., Sela,S., Bercovier,H., Pitulle,C. and
stackebrandt,E.** (1991): Complete nucleotide sequence of the
Mycobacterium leprae 23S and 5S rRNA genes plus flanking
regions and their potential in designing diagnostic
oligonucleotide probes. *FEBS Letters* **281**, 114-118.

Lindahl,L. and Zengel,J.M. (1986): Ribosomal genes in
Escherichia coli. *Annual Review of Genetics* **20**, 297-326.

Lindquist,S. and Craig,E.A. (1988): The heat shock
proteins. *Annual Review of Genetics* **22**, 631-677.

Liu-Johnson,H.N., Gartenberg,M.R. and Crothers,D.M. (1986):
The DNA binding domain and bending angle of E.coli CAP
protein. *Cell* **47**, 995-1005.

Loughney,K., Lund,E. and Dahlberg,J.E. (1982): tRNA genes
are found between the 16S and 23S rRNA genes in *Bacillus*
subtilis. *Nucleic Acids Research* **10**, 1607-1624.

Loughney,K., Lund,E. and Dahlberg,J.E. (1983): Ribosomal
RNA precursors of *Bacillus subtilis*. *Nucleic Acids Research*
11, 6709-6721.

Lund,E. and Dahlberg,J.E. (1976): Transfer RNA genes
between 16S and 23S rRNA genes in rRNA transcription units
of *E.coli*. *Cell* **7**, 165-177.

Lund.E. and Dahlberg,J.E. (1977): Spacer transfer RNAs in ribosomal RNA transcripts of *E.coli*: processing of 30S ribosomal RNA in vitro. *Cell* **11**, 247-262.

Mackie,T.J. and McCartney,J.E. (1949): Handbook of Practical Bacteriology. 11th ed. E & S Livingstone, Edinburgh.

Mason,S., Li,J. and Greenblatt,J. (1992): Host factor requirements for processive antitermination of transcription and suppression of pausing by the N protein of bacteriophage λ . *Journal of Biological Chemistry* **267** (27), 19418-19426.

Matsuo,K., Yamaguchi,R., Yamazaki,A., Tasaka,H. and Yamada,T. (1988): Cloning and expression of the *Mycobacterium bovis* BCG gene for extracellular alpha antigen. *Journal of Bacteriology* **170**, 3847-3854.

McClure,W.R. (1985): Mechanism and control of transcription in prokaryotes. *Annual Review of Biochemistry* **54**, 171-204.

McFadden,J.J., Butcher,P.D., Thompson,J., Chiodini,R. and Hermon-Taylor,J. (1987): The use of DNA probes identifying restriction-fragment-length polymorphisms to examine the *Mycobacterium avium* complex. *Molecular Microbiology* **1**, 283-291.

Mehra,V., Sweetser,D. and Young,R.A. (1986): Efficient mapping of protein antigenic determinants. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 7013-7017.

Meissner,G. and Schröder,K.H. (1975): Relationship between *Mycobacterium simiae* and *Mycobacterium habana*. *American Review of Respiratory Disease* **111**, 196-200.

Merkal,R.S. (1984): Paratuberculosis. In: *The mycobacteria. A source-book*, part B. (Eds: Kubica,GP; Wayne,LG) Marcel Dekker, Inc. New York, 1237-1249.

Miura,A., Krueger,J.H., Itoh,S., de Boer,H.A. and Nomura,M. (1981): Growth-rate-dependent regulation of ribosome synthesis in *E.coli*: expression of the *lacZ* and *galK* genes fused to ribosomal promoters. *Cell* **25**, 773-782.

Mizushima-Sugano,J. and Kaziro,Y. (1985): Regulation of the expression of the *tufB* operon: DNA sequences directly involved in the stringent control. *European Molecular Biology Organization Journal* **4**, 1053-1058.

Morgan,E.A. (1980): Insertions of Tn10 into an *E.coli* ribosomal RNA operon are incompletely polar. *Cell* **21**, 257-265.

Morgan,E.A. (1982): Ribosomal RNA genes in *Escherichia*

coli. Cell Nucleus **10**, 1-29.

Morgan, E.A. (1986): Antitermination mechanisms in rRNA operons of *Escherichia coli*. *Journal of Bacteriology* **168**, 1-5.

Nerland, A.H., Mustafa, A.S., Sweetser, D., Godal, T. and Young, R.A. (1988): A protein antigen of *Mycobacterium leprae* is related to a family of small heat shock proteins. *Journal of Bacteriology* **170**, 5919-5921.

Nodwell, J.R. and Greenblatt, J. (1993): Recognition of boxA antiterminator RNA by the E.coli antitermination factors NusB and ribosomal protein S10. *Cell* **72**, 261-268.

Noller, H.F. (1984): Structure of ribosomal RNA. *Annual Review of Biochemistry* **53**, 119-162.

Nomura, M., Gourse, R. and Baughman, G. (1984): Regulation of the synthesis of ribosomes and ribosomal components. *Annual Review of Biochemistry* **53**, 75-118.

Noordeen, S.K., Lopez-Bravo, L. and Sundaresan, T.K. (1992): Estimated numbers of leprosy cases in the world. *Leprosy Review* **63**, 282-287.

Norrande, J., Kempe, T. and Messing, J. (1983): Construction of improved M13 vectors using oligodeoxynucleotide directed

mutagenesis. *Gene* **26**, 101-106.

Ogasawara,N., Moriya,S. and Yoshikawa,H. (1983): Structure and organization of rRNA operons in the region of the replication origin of the *Bacillus subtilis* chromosome. *Nucleic Acids Research* **11**, 6301-6318.

Ota,Y., Kikuchi,A. and Cashel,M. (1979): Gene expression of an *Escherichia coli* ribosomal RNA promoter fused to structural genes of the galactose operon. *Proceedings of the National Academy of Sciences of the United States of America* **76**, 5799-5803.

Patel,B.K.R., Banerjee,D.K. and Butcher,P.D. (1991): Extraction and characterization of mRNA from mycobacteria: implication for virulence gene identification. *Journal of Microbiological Methods* **13**, 99-111.

Pernodet,J.L., Boocard,F., Alegre,M.T., Gagnat,J. and Guerineau,M. (1989): Organization and nucleotide sequence analysis of a ribosomal RNA cluster from *Streptomyces ambofaciens*. *Gene* **59**, 33-46.

Rado,T.A., Bates,J.H. and Fitzhugh,J.K. (1976): Evidence for host-dependant modification and restriction of bacteriophage DNA in *Mycobacterium tuberculosis*. *Journal of General Virology* **30**, 91-97.

Raue,H.A., Musters,W., Rutgers,C.A., Riet,J.V. and Planta,R.J. (1990): In: *The ribosome, structure, function, and evolution*. (Ed: Hill et al.) American Society for Microbiology, Washington,DC, 217-235.

Rogall,T., Flohr,T. and Böttger,E.C. (1990a): Differentiation of *Mycobacterium* species by direct sequencing of amplified DNA. *Journal of General Microbiology* **136**, 1915-1920.

Rogall,T., Wolters,J., Flohr,T. and Böttger,E.C. (1990b): Towards a phylogeny and definition of species at the molecular level within the genus *Mycobacterium*. *International Journal of Systemic Bacteriology* **40**, 323-330.

Roy,M.K., Singh,B., Ray,B.K. and Apirion,D. (1983): Maturation of 5S rRNA: Ribonuclease E cleavages and their dependence on precursor sequences. *European Journal of Biochemistry* **131**, 119-127.

Ryals,J., Little,R. and Bremer,H. (1982a): Control of RNA synthesis in *Escherichia coli* after a shift to higher temperature. *Journal of Bacteriology* **151**, 1425-1432.

Ryals,J., Little,R. and Bremer,H. (1982b): Control of rRNA and tRNA syntheses in *Escherichia coli* by guanosine tetraphosphate. *Journal of Bacteriology* **151**, 1261-1268.

Saiki,R.K., Gelfand,D.H., Stoffel,S., Sharf,S.J., Higuchi,R., Horn,G.T., Mullis,K.B. and Erlich,H.A. (1988): Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487.

Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989): Molecular Cloning, a Laboratory Manual. 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Sarmientos,P., Svlvester,J.E., Contente,S. and Cashel,M. (1983): Differential stringent control of the tandem *E.coli* ribosomal RNA promoters from the *rrnA* operon expressed in vivo in multicopy plasmids. *Cell* **32**, 1337-1346.

Sawada,M., Muto,A., Iwami,M., Yamao,F. and Osawa,S. (1984): Organization of ribosomal RNA genes in *Mycoplasma capricolum*. *Molecular and General Genetics* **196**, 311-316.

Schlessinger,D. (1980): Processing of ribosomal RNA transcripts in bacteria. In: *Ribosomes*. (Eds: Chambiliss,G; Craven,GR; Davies,K; Khan,L; Nomura,M) Baltimore: University Park Press. 767-780.

Sela,S. and Clark-Curtiss,J.E. (1991): Cloning and characterization of the *Mycobacterium leprae* putative ribosomal RNA promoter in *Escherichia coli*. *Gene* **98**, 123-127.

Sela,S., Clark-Curtiss,J.E. and Bercovier,H. (1989): Characterization and taxonomic implications of the rRNA genes of *Mycobacterium leprae*. *Journal of Bacteriology* **171** (1), 70-73.

Shankar,S. and Tyagi,A.K. (1992a): MfoAI, a novel isoschizomer of HaeIII from *Mycobacterium fortuitum* recognizing 5'-GG/CC-3'. *Nucleic Acids Research* **20**, 2890.

Shankar,S. and Tyagi,A.K. (1992b): MhaAI, a novel isoschizomer of PstI from *Mycobacterium habana* recognizing 5'-CTGCA/G-3'. *Nucleic Acids Research* **20**, 2891.

Shen,W.F., Squires,C. and Squires,C.L. (1982): Nucleotide sequence of the *rrnG* ribosomal RNA promoter region of *Escherichia coli*. *Nucleic Acids Research* **10**, 3303-3313.

Shine,J. and Dalgarno,L. (1974): The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proceedings of the National Academy of Sciences of the United States of America* **71**, 1342-1346.

Sirakova,T.D., Bardarov,S.S., Kriakov,J.I. and Kriakov,K.I. (1989): Molecular cloning of mycobacterial promoters in *Escherichia coli*. *FEMS Microbiology Letters* **59**, 153-156.

Snapper,S.B., Lugosi,L., Jekkel,A., Melton,R., Kieser,T.,

Bloom,B.R. and Jacobs,W.R. (1988): Lysogeny and transformation in mycobacteria: stable expression of foreign genes. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 6987-6991.

Snapper,S.B., Melton,R.E., Mustafa,S., Kieser,T. and Jacobs Jr,W.R. (1990): Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Molecular Microbiology* **4**, 1911-1919.

Sommers,H.M. and Good,R.C. (1985): *Mycobacterium*. In: Manual of clinical microbiology. 4th ed. (Eds: Lennetee,EH; Balows,A; Hausler,Jr WJ; Shadomy,HJ) American Society for Microbiology, Washington,D.C. 216-248.

Stackebrandt,E. and Woese,C.R. (1981): The evolution of prokaryotes. *Symposia of the Society for General Microbiology* **32**, 1-31.

Steege,D.A., Cone,K.C., Queen,C. and Rosenberg,M. (1987): Bacteriophage λ N gene leader RNA. *Journal of Biological Chemistry* **262**, 17651-17658.

Stewart,G.C. and Bott,K.F. (1983): DNA sequence of the tandem ribosomal RNA promoter for *B.subtilis* operon *rrnB*. *Nucleic Acids Research* **11**, 6289-6300.

Suzuki,Y., Nagata,A., Ono,Y. and Yamada,I. (1988): Complete

nucleotide sequence of the 16S rRNA gene of *Mycobacterium bovis* BCG. *Journal of Bacteriology* **170**, 2886-2889.

Taschke, C. and Herrmann, R. (1986): Analysis of transcription and processing signals of the 16S-23S rRNA operon of *Mycoplasma hyopneumoniae*. *Molecular and General Genetics* **205**, 434-441.

Teske, A., Wolters, J. and Böttger, E.C. (1991): The 16S rRNA nucleotide sequence of *Mycobacterium leprae*: phylogenetic position and development of DNA probes. *FEMS Microbiology Letters* **80**, 231-238.

Thole, J.R., Dauwese, H.G., Das, P.K., Groothuis, D.G., Schouls, L.M. and van Embden, J.D.A. (1985): Cloning of *Mycobacterium bovis* BCG DNA and expression of antigens in *Escherichia coli*. *Infection and Immunity* **50**, 800-806.

Travers, A.A. (1980a): Promoter sequence for stringent control of bacterial ribonucleic acid synthesis. *Journal of Bacteriology* **141**(2), 973-976.

Travers, A.A. (1980b): A tRNA^{tyr} promoter with an altered *in vitro* response to ppGpp. *Journal of Molecular Biology* **141**, 91-97.

Travers, A.A., Lamond, A.I. and Mace, H.A.F. (1982): ppGpp Regulates the binding of two RNA polymerase molecules to

the *tyrT* promoter. *Nucleic Acids Research* **10**, 5043-5057.

Travers,A.A., Lamond,A.I. and Weeks,J.R. (1986): Alteration of the growth rate dependent regulation of *E.coli tyrT* expression by promoter mutations. *Journal of Molecular Biology* **189**, 251-255.

Watson,N. and Apirion,D. (1981): Ribonuclease F, a putative processing endoribonuclease from *Escherichia coli*. *Biochemical and Biophysical Research Communications* **103**, 543-551.

Wawrousek,E.F. and Hansen,N.J. (1983): Structure and organization of a cluster of six tRNA gene sets in *Bacillus subtilis*. *Journal of Biological Chemistry* **258**, 291-298.

Weiss,R.B., Dunn,D.M., Atkins,J.F. and Gesteland,R.F. (1987): *Cold Spring Harbor Symposium Quantitative Biology* **52**, 687-693.

Wezel,G.P.V., Vijgenboom,E. and Bosch,L. (1991): A comparative study of the ribosomal RNA operons of *Streptomyces coelicolor* A3(2) and sequence analysis of *rrnA*. *Nucleic Acids Research* **19**, 4399-4403.

Winder,F.G. and Rooney,S.A. (1970): Effects of nitrogenous components of the medium on the carbohydrate and nucleic acid content of *Mycobacterium tuberculosis* BCG. *Journal of*

General Microbiology **63**, 29-39.

Woese, C.R. (1987): Bacterial evolution. *Microbiological Reviews* **51**, 221-271.

Woese, C.R., Gutell, R., Gupta, R. and Noller, H.F. (1983): Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiological Reviews* **47**, 621-669.

Yamaguchi, R., Matsuo, K., Yamazaki, A., Abe, C., Nagai, S., Terasaka, K. and Yamada, T. (1989): Cloning and characterization of the gene for immunogenic protein MPB64 of *Mycobacterium bovis* BCG. *Infection and Immunity* **57**, 283-288.

Young, D.B., Lathigra, R., Hendrix, R., Sweetser, D. and Young, R.A. (1988): Stress proteins are major immune targets in leprosy and tuberculosis. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 4267-4270.

Young, R.A., Bloom, B.R., Grosskinsky, C.M., Ivanyi, J., Thomas, D. and Davis, R.W. (1985b): Dissection of *M.tuberculosis* antigens using recombinant DNA. *Proceedings of the National Academy of Sciences of the United States of America* **82**, 2583-2587.

Young,R.A., Macklis,R. and Steitz,J.A. (1979): Sequence of the 16S-23S spacer region in two ribosomal RNA operons of *Escherichia coli*. *Journal of Biological Chemistry* **254**(9), 3264-3271.

Young,R.A., Mehra,V., Sweetser,D., Buchanan,T., Clark-Curtiss,J., Davis,R.W. and Bloom,B.R. (1985a): Genes for the major protein antigens of the leprosy parasite *M.leprae*. *Nature* **316**, 450-452.

Young,R.A. and Steitz,J.A. (1978): Complementary sequences 1700 nucleotides apart form a ribonuclease III cleavage site in *Escherichia coli* ribosomal precursor RNA. *Proceedings of the National Academy of Sciences of the United States of America* **75**, 3503-3507.

Zaug,A.J., Been,M.D. and Cech,T.R. (1986): The *Tetrahymena* ribozyme acts like an RNA restriction endonuclease. *Nature* **324**, 429-433.